

Interdisciplinary modelling and analysis of infectious diseases

Cathal Mills

Department of Statistics

University of Oxford

Supervised by Professor Christl Donnelly, Professor Moritz Kraemer, and Professor Ben Lambert

DPhil in Statistics
June 2025

"I believe I have a personal responsibility to make a positive impact on society."

Anthony Fauci, 2005 [1]

“When a law is right it can be used to find another one. If we have confidence in a law, then if something appears to be wrong it can suggest to us another phenomenon.”

Richard Feynman, 1965 [2]

Abstract

Infectious diseases pose substantial and continuous threats to global health. From the Black Death to the influenza pandemics, and more recently COVID-19, the impacts of infectious disease outbreaks have long highlighted the need for a quantitative understanding of epidemic dynamics. Epidemiological modelling is about providing such an understanding. As data are always imperfect, models are also imperfect, yet models remain useful tools that create manageable abstractions of reality to quantitatively describe why, when, and where infectious diseases spread. These modelling objectives are grounded in practical aims – to inform policy, advance scientific research, and benefit public health.

Public health authorities face considerable uncertainty and time pressures when making difficult and consequential decisions about epidemic control. This necessitates timely and reliable information about the past, current, and likely future epidemic dynamics in order to respond appropriately with targeted strategies for short-term outbreak responses and longer-term planning. As argued in the current thesis, the power of our modelling lies in the ability to provide such information, developing methods that can i) learn from the past, ii) monitor the present, and iii) plan for the future.

Despite an increasing adoption of multi-model and interdisciplinary approaches, epidemiological modelling can sometimes be hampered by narrow perspectives that isolate models, data, or branches of science. Here, we will describe how we can often improve upon restrictive modelling approaches, even in resource-limited settings. We will i) allow the research questions to determine the methodology and ii) reach beyond single modelling disciplines, data streams, and branches of science to answer the questions about epidemic dynamics and drivers. While still dependent on data collection efforts and well-developed, public-health-relevant models, we aim to show that this multi-model, interdisciplinary perspective can often provide a greater volume and variety of insights to reliably inform public health policy and hopefully benefit society.

The COVID-19 pandemic, caused by the spread of SARS-CoV-2, had devastating consequences in societies globally. The urgent need to understand and control the spread led to a large volume and variety of data streams being available for monitoring epidemic dynamics and drivers. Two such data streams came from wastewater-based epidemiology (WBE) and community infection prevalence surveys, which offered potential avenues to complement costly and typically incomplete disease case surveillance. In Chapter 3, we performed the first nationwide analysis of the utility of WBE for monitoring community prevalence of SARS-CoV-2 at high spatial resolution throughout a nation across the COVID-19 pandemic. Introducing a new geospatial mapping approach and shedding metric, we found that i) higher vaccination coverage, rising immunity levels, and Omicron SARS-CoV-2 variants all induced lower faecal shedding over time and ii) WBE could complement, but not replace, prevalence surveys.

The main focus of the thesis is dengue, the world's most widespread and fastest expanding mosquito-borne disease. Dengue virus transmission is driven by the coalescence of environmental, human, and biological influences. In Chapter 4, we introduced a multi-model workflow that integrates heterogeneous data streams in an approach that unifies and extends techniques for climate-informed modelling and forecasting of dengue epidemics. Applying our workflow to dengue epidemics in northern Peru, we used wavelets and Bayesian hierarchical models to quantify both spatiotemporal dynamics and climatic drivers of epidemics at varying temporal and spatial resolutions. Our ensemble models for forecasting also blended the strengths of different modelling disciplines to produce statistically rigorous and public-health-focused outputs. Continuing this theme of multi-model approaches, in Chapter 5, we investigated the relative importance of human movement in shaping urban dengue outbreaks in the neighbourhoods of Fortaleza, Brazil. Here, our new evidence synthesis approach melded Bayesian hierarchical and deep learning models with new theory, which allowed us to unveil a consistently high importance of human movement within and across dengue seasons and a relatively constant spatial profile for source-sink dynamics of dengue epidemics in neighbourhoods. Adopting a mechanistic modelling perspective, in Chapter 6, we worked to address the lack of methods for monitoring how fast vector-borne pathogens are spreading over time. To this end, we introduced new theory to extend the popular renewal equations that form the basis of most methods for estimating

time-varying reproduction numbers, $R(t)$, for directly transmitted diseases. Starting from age-structured models of dengue transmission across humans and mosquitoes, we derived renewal equations that control how past human infections generate new human infections. This work provides theoretical foundations for inference frameworks that estimate $R(t)$ for vector-borne pathogens and, ultimately, inform epidemic control strategies. To better understand how to simulate and parameterise these models, in Chapter 7, we developed new theory that extends the seminal Gillespie algorithms to exactly simulate stochastic systems which experience external variability and time-varying propensities between events. These algorithms, illustrated for a model of dengue transmission, are generalisable to simulating stochastic reaction processes that are used across many branches of science, from chemistry to biology.

Chapter 8 captures the simultaneous theoretical and public health focus of this thesis, as we provide a new evaluation framework for forecasters and forecast evaluators. This framework shifts the perspective of forecast evaluation to the decision-maker. Translating metrics and ideas from many branches of science to the context of epidemics, we provide methods to measure the value of forecasts by their ability to inform the difficult and uncertain epidemic-control-related decisions faced by public health authorities. This evaluation framework connects statistical concepts with principles from decision theory and information theory to promote more useful and reliable infectious disease forecasts for improved public health responses. In Chapter 9, we briefly summarise two published manuscripts: i) our research on the future of artificial intelligence for modelling infectious diseases, and ii) our systematic review of literature from the COVID-19 pandemic on the effectiveness of social distancing measures for epidemic control. Finally, in Chapter 10, we summarise our methods and analyses and discuss how these address our stated objectives, how they fit in the wider epidemiological context, and how they could be extended to address limitations and explore new policy and research questions in resource-limited settings.

As public health authorities combat persistent infectious disease threats, tackle the related challenges of climate change and conspiracy theories, and prepare for future epidemics and pandemics, our interdisciplinary methods will support authorities by providing robust, timely, and actionable information to improve global public health.

Acknowledgements

Dear all, thank you.

There is a long list of people to thank for supporting me, and my words will inevitably not do justice to the gratitude I feel. In any case, much like my thesis, I thought that I should at least try.

Christl, Moritz, and Ben – Perhaps fittingly, it is incredibly difficult to quantify the contributions that each of you has each made. I could not have asked for a more knowledgeable, supportive, and encouraging team of mentors. It is a privilege to be able to work so closely with such skilled scientists and, more importantly, incredible people. I am already looking forward to our next chapter together.

First, to Christl, it is not hyperbolic to state that I have no idea where I would be, or what I would be doing, if you had not embraced my naïve request to work together. You provided me with the opportunity to join a team and institution of extraordinary researchers, paved the paths for my many research endeavours, and most importantly, provided the care, humour, and support that I value immensely. Your unwavering dedication to your students, while being Head of Department, has long been acknowledged by others, yet I cannot exaggerate how remarkable this is. I have learned so much from you and will forever be indebted to you, not just in free food or proof-reading.

To Ben and Moritz, thank you for providing the most useful Transfer of Status. Ben – Thank you for always listening, thinking deeply, and caring so much. Since I arrived at your door with a promisingly incorrect idea, we have had great fun (and seemingly some success based on the thesis) as we made some complex things slightly less complex. Beyond research, I am deeply grateful for you always teaching me new things, embracing my bad humour, and supporting me so closely in my next steps. Your attention to detail in everything we have done together is something I can only hope to replicate in my own research. Moritz – You welcomed me into an incredible group, provided opportunities that I will forever cherish (both research- and travel-related), and you have always been a supportive voice to listen to my latest (usually) ridiculous idea. Your relentless passion for science and wide-ranging knowledge have been crucial to my development and enjoyment of my DPhil. I can only hope that our research together is just the beginning of a long journey. It is a testament to both of you (Moritz and Ben) that I am not alone in staying around to collaborate with you, or perhaps just annoy you with exciting ideas.

To Olli, thank you for being such an amazing MSc supervisor and lecturer and now collaborator. As a researcher, your insights are always on point, while your friendliness and encouragement of my research will forever be cherished by me. I am so lucky to still have the opportunity to work with you on our many shared interests.

To Team Donnelly, in particular Sarah, Ruth, Nic, Matt, and Tarek, I could not have asked for a more welcoming, funny, and wild group of friends to share the DPhil journey with. Thank you once again to Christl for creating this truly special and unique environment. It is an honour to call each of you a colleague and a dear friend. To Sarah, thank you for being a friendly, welcoming, and eternally positive colleague. I will forever remember our train odyssey to Bologna, our bonding over climate change, and your relentless passion for our social gatherings. To Ruth, thank you for being my DPhil academic idol, for always being there to talk, listen, and laugh, and for embracing our brilliantly chaotic Friday meetings. We still feel your absence. To Nic, thank you for always providing humour, expertise, and scepticism in equal measure. I have learned so much from you and will deeply miss our seriously unserious and unseriously serious conversations. To Matt, thank you for being so witty, friendly, and mathematically inspiring. I miss your uniquely cheerful presence in the office and it was incredible to have someone to discuss both maths and football. To Tarek, thank you for seamlessly integrating into and contributing so much to our weird but wonderful group and for always being my brilliant travel companion and organiser. I greatly value both your humour and insights, and our friendship. To Alex, Ben R., Junjie, and Younjung, it has been a pleasure to get to know you over the past few months. I cannot wait to get to know each of you better and work with equally amazing scientists.

There is a longer list of excellent researchers and collaborators who, at various stages, provided much-needed words of encouragement that made me feel like less of a research imposter. In no order (well, I used an RNG), a special thank you to Robin, Sen, Sumali, Charlie, Adam, Bernardo, Rosie, Ilaria, Will, Joseph, Kris, and Claudia. I am so fortunate to work with each of you.

To all my special friends, from childhood to Oxford. To my childhood friends, particularly Leon, Conúil, John, and Jack, you may never read this but thank you for the many great times we had together. To those in Oxford, thank you to Vik, Max, Isobel, and Laura for welcoming a unusually sounding, fast-talking guy into a great office in a weird department. To Zoi, thank you for always being there to keep the office warm, never talk about research, and provide much-needed company, friendship and humour in more difficult times. To Dan, thank you for being the best and most humorous flatmate, for ignoring my many quirks, and for always being there to listen and/or laugh. To Chris, Jakiw, Silvia, Stefano, and Tyler, thank you for being friends who could always provide the social appetite and energy that I most often lacked.

To Séamus, I could not have asked to have grown up (metaphorically at least, not physically) with a more caring, supportive, and brilliant brother. Thank you for always embracing my nerdiness, coping with my many faults, and being my best friend.

To my aunts, uncles, and cousins in the east and west of Ireland and beyond, thank you for never doubting me and always encouraging me.

To my parents, the following words will understate my gratitude. Thank you for *everything*. To my Mum, thank you for sharing your love for maths, the Irish language and football, for being my biggest cheerleader in everything that I do, and for always being the closest person to me in my life. To my Dad, thank you for always being there for me; for supporting me, for being eternally positive, and for providing humour, wit, and care every day of my life.

I dedicate this thesis to the uncles that I never met and to the grandparents whose struggles, fight, and hard work have made my journey possible. I miss you all.

Contents

1 Introduction **1**

1.1 DPhil thesis background . . . 1
1.2 DPhil thesis aims and objectives . . . 2
1.3 DPhil thesis structure . . . 4

2 Background and literature review **6**

2.1 Epidemiology of COVID-19 and dengue . . . 6
2.1.1 COVID-19 . . . 6
2.1.2 Dengue . . . 7
2.2 Fundamental question: *Why* model infectious disease outbreaks? . . . 9
2.3 Epidemiological data and derived quantities . . . 10
2.3.1 Informing early outbreak response . . . 10
2.3.2 Measuring time-varying epidemic properties . . . 13
2.3.3 Predicting epidemic trajectories . . . 15
2.4 Challenges for epidemiological analyses . . . 16
2.4.1 Reporting rates, representativeness, and standardised data . . . 16
2.4.2 Delayed and truncated data . . . 16
2.4.3 Averages, heterogeneities, and communication . . . 17
2.4.4 Nonlinearities . . . 18
2.4.5 Predictions, predictability, and governing laws . . . 18
2.4.6 Spatiotemporal obstacles . . . 19
2.5 Complementary data sources . . . 20
2.5.1 Human mobility data . . . 21
2.5.2 Climate and weather data . . . 22
2.5.3 More novel data streams . . . 23
2.5.4 Seroprevalence data . . . 23
2.5.5 Raw data versus derived quantities . . . 24
2.6 Modelling infectious disease outbreaks . . . 24
2.6.1 Mechanistic models . . . 26
2.6.2 Statistical models . . . 28
2.6.3 AI: Machine learning and deep learning models . . . 31
2.6.4 Interdisciplinary approaches . . . 33
2.6.5 Simulation, inference and optimisation . . . 35
2.7 Model evaluation and evidence synthesis . . . 37
2.7.1 Model evaluation: Information criteria and scoring rules . . . 38
2.7.2 Multiple model comparison, evidence synthesis, and communication . . . 39

3 The utility of wastewater surveillance for monitoring SARS-CoV-2 prevalence **40**

3.1 Introduction . . . 41
3.2 Results . . . 41
3.2.1 Relationship Between SARS-CoV-2 RNA Wastewater Concentrations and Prevalence . . . 43
3.2.2 Out-of-Sample Wastewater-Model-Based Estimates of REACT-1 Prevalence . . . 44
3.2.3 Complementary Use of WBE for Monitoring SARS-CoV-2 Prevalence . . . 46

3.3 Discussion . . . 48
3.4 Materials and Methods . . . 51
3.4.1 Study Design . . . 51
3.4.2 Geospatial Mapping of Wastewater Concentrations and Relating to REACT-1 . . . 52
3.4.3 Relationship Between Wastewater Concentrations and REACT-1 Prevalence . . . 53
3.4.4 Modelling REACT-1 SARS-CoV-2 Prevalence . . . 53
3.4.5 Types of Wastewater-Model-Based Estimates . . . 54
3.4.6 Data Availability . . . 55

4 Unifying techniques for climate-informed dengue modelling and forecasting 56
4.1 Paper I: Climate-based modelling and forecasting of dengue in three endemic departments of Peru . . . 56
4.1.1 Introduction . . . 57
4.1.2 Methods . . . 58
4.1.3 Results . . . 62
4.1.4 Discussion . . . 65
4.2 Paper II: Multi-model approach to understand and predict past and future dengue epidemic dynamics . . . 69
4.2.1 Introduction . . . 69
4.2.2 Materials and methods . . . 71
4.2.3 Results . . . 76
4.2.4 Discussion . . . 80

5 The time- and space-varying roles of human mobility in shaping urban dengue outbreaks 87
5.1 Introduction . . . 88
5.2 Materials and methods . . . 89
5.2.1 Study area and data . . . 89
5.2.2 Model development and fitting periods . . . 89
5.2.3 Bayesian hierarchical models . . . 90
5.2.4 Data-driven, deep learning models . . . 92
5.2.5 Evidence synthesis . . . 94
5.3 Results . . . 94
5.3.1 Mobility consistently drives urban dengue outbreaks . . . 95
5.3.2 Relative importance of human mobility over time to individual neighbourhoods . . . 95
5.3.3 Space-varying human-mobility-driven influences of neighbourhoods . . . 97
5.4 Discussion . . . 99

6 Reproduction numbers for vector-borne diseases 104
6.1 Introduction . . . 105
6.2 Methods and results . . . 106
6.2.1 Infection-age-structured human and mosquito populations . . . 106
6.2.2 Exposure-and-infection-age-structured populations of human and mosquitoes . . . 112
6.3 Discussion . . . 116

7 Exact simulation of stochastic reaction processes with external variability 119
7.1 Introduction . . . 120
7.2 The classical Gillespie simulation algorithms for fixed propensities between events . . . 122
7.2.1 Gillespie's direct method . . . 122
7.2.2 Gillespie's first reaction method . . . 123
7.3 Gillespie simulation algorithms for time-varying propensities between events . . . 125
7.3.1 Gillespie's direct method . . . 125
7.3.2 Gillespie's first reaction method . . . 128
7.4 Applications of algorithms to chemical reaction systems with time-varying propensities . . . 129
7.4.1 Problem setup . . . 129
7.4.2 Numerical ODE simulation . . . 130
7.4.3 Time-varying birth and death rates and stochastic simulation details . . . 130
7.4.4 Results of applications . . . 132
7.5 Individual-based model of dengue transmission . . . 132

7.5.1 Problem setup . . . 132
7.5.2 Application details and results . . . 135
7.6 Discussion . . . 136

8 From metric to action: An evaluation framework to translate infectious disease forecasts into policy decisions 139
8.1 Introduction . . . 140
8.2 Recasting scoring metrics for interpretable epidemic forecasting across space and time . . . 141
8.2.1 Scoring functions and rules . . . 141
8.2.2 Stratifications and interpretable scoring metrics . . . 143
8.2.3 Elementary scores and utility functions . . . 144
8.2.4 Eliciting risk preferences for decision-makers . . . 146
8.3 Forecast value for informing epidemic decision-making . . . 147
8.3.1 User-and-event-focused scoring metrics . . . 147
8.4 Risk-averse epidemic decision-making and distributionally robust optimisation . . . 150
8.5 Predictability assessments for epidemic forecasting . . . 150
8.5.1 Introducing predictability for infectious disease forecasting . . . 151
8.5.2 Permutation entropy . . . 151
8.5.3 Embedding predictability in our framework . . . 152
8.5.4 Extending predictability analyses . . . 152
8.6 Applying metrics to epidemic forecasts . . . 153
8.6.1 Application of our framework . . . 153
8.6.2 From metric to action I: Evaluation workflow results . . . 153
8.6.3 From metric to action II: Real-time decision-making . . . 157
8.7 Discussion . . . 159

9 Other publications 163
9.1 Artificial Intelligence for Infectious Disease Epidemiology . . . 163
9.2 Systematic review of the effectiveness of social distancing measures and lockdowns during the COVID-19 pandemic . . . 164

10 Discussion and conclusion 165
10.1 Learning from the past: Relative roles of epidemic drivers . . . 165
10.1.1 Coalescing human, biological, and environmental drivers of dengue . . . 165
10.2 Understanding the present: Monitoring epidemic dynamics . . . 173
10.2.1 Reproduction numbers for mosquito-borne diseases . . . 173
10.2.2 Wastewater-based epidemiology (WBE) . . . 181
10.3 Planning for the future: Forecasting epidemic trajectories . . . 183
10.3.1 Forecasting models . . . 183
10.3.2 Forecast evaluation . . . 189
10.4 Conclusion . . . 191

References 192

A Appendix — The utility of wastewater surveillance for monitoring SARS-CoV-2 prevalence i
A.1 Relationship Between SARS-CoV-2 RNA Wastewater Concentrations and Prevalence . . . i
A.2 Wastewater-Based Model Covariates . . . iv
A.3 Out-of-Sample Wastewater-Model-Based Estimates of Prevalence . . . v
A.4 Complementary Use of WBE . . . xii
A.5 Population-level faecal shedding . . . xiii
A.6 Geospatial Population Estimates (GPEs) . . . xiv
A.7 Further Details of REACT-1 . . . xvi
A.8 Wastewater Sources of Uncertainty . . . xvi
A.9 Sensitivity of Geospatial Mapping Approach . . . xvii
A.10 Bayesian Hierarchical Modelling . . . xviii
A.11 Vaccination-Log Concentration Interaction Variable . . . xx
A.12 Acronyms . . . xxi

A.13 Timeline of REACT-1 and Estimated Number of LTLAs in EMHP Programme . . . xxii
A.14 Wastewater Variant Detections . . . xxiii
A.15 Further Results . . . xxiv

B Appendix — Climate-based modelling and forecasting of dengue in three endemic departments of Peru xxv
B.1 Acronyms . . . xxvi
B.2 Further study area details . . . xxvii
B.3 Further model framework details . . . xxvii
B.3.1 Population Offset . . . xxvii
B.3.2 Fixed effects . . . xxvii
B.3.3 Temporal random effects . . . xxviii
B.3.4 Spatiotemporal random effects . . . xxviii
B.3.5 Specification of climate-based DLNMs . . . xxviii
B.4 Figures . . . xxx
B.4.1 Exploratory analyses . . . xxx
B.4.2 Sensitivity analyses . . . xxxi
B.4.3 Spatiotemporal climate trends . . . xxxiv
B.4.4 Seasonality in dengue incidence rates . . . xxxiv
B.4.5 Cumulative risk induced by climatic variables . . . xxxv
B.4.6 Retrospective modelling in-sample predictive fit . . . xxxv
B.4.7 Estimated yearly random effects by department . . . xxxvi
B.4.8 Estimated monthly random effects by department . . . xxxvi
B.4.9 Estimated spatially structured random effects by department . . . xxxvii
B.4.10 Estimated effect of momentum indicator . . . xxxviii
B.4.11 Visualisation of climate-based out-of-sample errors . . . xxxviii
B.4.12 Reliability of retrospective modelling outbreak detection . . . xl
B.4.13 Visualisation of climate-based forecasting errors . . . xli
B.4.14 Reliability of forecasting outbreak detection . . . xli
B.4.15 Non-model-based simple forecasting outbreak detection . . . xlii

C Appendix — Multi-model approach to understand and predict past and future dengue epidemic dynamics xliii
C.1 Acronyms . . . xliv
C.2 Ensemble model abbreviations . . . xlv
C.2.1 Forecasting models . . . xlv
C.3 EPIFORGE 2020 guidelines . . . xlvi
C.4 Climate-based model: Additional details . . . xlvii
C.4.1 Delayed nonlinear climatic effects . . . xlvii
C.4.2 Temporal and spatial random Effects . . . xlvii
C.4.3 Fixed Effects . . . xlviii
C.4.4 Covariates . . . xlix
C.5 Ensemble framework: Additional details . . . l
C.5.1 Background to probabilistic ensemble forecasting . . . l
C.5.2 Trained and untrained ensemble frameworks . . . l
C.6 Ensemble component forecasters: Additional details . . . l
C.7 Outbreak detection: Additional details . . . li
C.7.1 Classification of DIR exceeding thresholds . . . li
C.8 Additional discussion of limitations . . . li
C.9 Figures and tables . . . liii
C.9.1 Wavelet and climate-based analyses . . . liii
C.9.2 Climate-based modelling . . . lxix
C.9.3 Probabilistic ensemble forecasting . . . lxx
C.9.4 Additional figures . . . xcvi

D Appendix — The time- and space-varying roles of human mobility in shaping urban dengue outbreaks **xcvii**
D.0.1 Notation and abbreviations . . . xcvii
D.1 Alternative candidate statistical models . . . xcix
D.1.1 Statistical model I: Space-varying mobility effects . . . xcix
D.1.2 Statistical model II: Space-varying mobility with autoregressive processes . . . xcix
D.1.3 Statistical model III: Time-varying mobility with neighbourhood autoregressive effects . . . ci
D.2 Results . . . cii
D.2.1 Plots of data . . . cii
D.2.2 Time-varying relative importance of human mobility . . . cv
D.2.3 Space-varying human-mobility-driven roles of neighbourhoods . . . cx

E Appendix — Reproduction numbers for vector-borne diseases **cxii**
E.1 Renewal equations, generation times, and McKendrick–von Foerster equations . . . cxii
E.2 Exposure-and-infection-age-structured human and infection-age-structured mosquito population . . . cxiv
E.2.1 Modelling framework . . . cxiv
E.2.2 Deriving a renewal equation and instantaneous reproduction number . . . cxiv
E.2.3 Relation to existing renewal equation frameworks . . . cxvi
E.2.4 Analytical solutions for R_0 and r . . . cxviii
E.3 Exposure-and-infection-age-structured human and infection-age-structured mosquito population . . . cxviii
E.3.1 Modelling framework . . . cxviii
E.3.2 Deriving a renewal equation and instantaneous reproduction number . . . cxix
E.3.3 Relation to the simple ubiquitous renewal equation frameworks . . . cxxii
E.4 Exposure-and-infection-age-structured populations of human and mosquitoes . . . cxxiv
E.4.1 Modelling framework . . . cxxiv
E.4.2 Deriving a renewal equation and time-varying reproduction number . . . cxxv
E.4.3 Between-stage reproduction numbers . . . cxxviii
E.4.4 Application to simulated and real-world data . . . cxxix

F Appendix — Exact simulation of stochastic reaction processes with external variability **cxxxiv**
F.1 Further results for simple chemical reaction systems . . . cxxxiv
F.1.1 Exponentially decreasing rates . . . cxxxiv
F.1.2 Linearly increasing rates . . . cxxxvi
F.2 Details and results of application to system of dengue transmission . . . cxxxviii
F.2.1 Symbols of application . . . cxxxviii
F.2.2 Parametric assumptions . . . cxxxix

G Appendix — From metric to action: An evaluation framework to translate infectious disease forecasts into policy decisions **cxl**
G.1 Extended table of concepts . . . cxli
G.2 Relative Economic Value (REV) . . . cxlii
G.3 EPIFORGE 2020 guidelines . . . cxliii
G.4 COVID-19 Forecast Hub Application . . . cxliv
G.5 Additional figures . . . cxlv
G.5.1 Weighted Interval Score (WIS) across space and time . . . cxlv
G.5.2 Relative Economic Value (REV) – Differing events and reference models . . . cli
G.5.3 Relative pinball loss (rPL) across space and time . . . clv
G.5.4 User-specific Murphy diagrams for higher C/L ratios . . . clxvi
G.5.5 Predictability versus forecast metrics . . . clxvii
G.5.6 Predictability and $R(t)$ . . . clxxiii
G.6 COVID-19 wave further analyses . . . clxxv
G.6.1 Pre-wave results . . . clxxv
G.6.2 Brier score decomposition before and during wave . . . clxxvi
G.6.3 Event-specific Murphy diagrams before and during wave . . . clxxvii
G.6.4 User-specific Murphy diagrams before and during wave . . . clxxviii
G.6.5 Pinball loss decomposition before and during wave . . . clxxix
G.6.6 Relative pinball loss before and during wave . . . clxxx
G.6.7 Predictability impacts before and during wave . . . clxxxi

G.6.8 Predictability versus forecast metrics clxxxii
G.7 Distributionally robust optimisation for risk-averse decision-making cxci

Acronyms

Acronym	Description
ABC	Approximate Bayesian Computation
ABM	Agent-Based Model
ACF	Autocorrelation Function
ADE	Antibody-Dependent Enhancement
Ae. aegypti	*Aedes aeqypti*
Ae. albopictus	*Aedes albopictus*
AI	Artificial Intelligence
AIC	Akaike Information Criterion
AUC	Area Under the Curve
BIC	Bayesian Information Criterion
BS	Brier Score
BYM	Besag, York and Mollié
CDC	Centre for Disease Control
CDF	Cumulative Distribution Function
CI	Confidence Interval
C/L	Cost-Loss
CLT	Central Limit Theorem
CMJ	Crump-Mode-Jagers
CORP	Consistent, Optimally binned, Reproducible, and Pool-adjacent-violators
CrI	Credible Interval
CRPS	Continuous Rank Probability Score
CTMC	Continuous Time Markov Chain
CV	Cross-Validated
DDT	Dichlorodiphenyltrichloroethane
DENV	Dengue Virus
DIC	Deviance Information Criterion
DIR	Dengue Incidence Rate
DL	Deep Learning
DLNM	Distributed Lag Nonlinear Models
DRO	Distributionally Robust Optimisation
DSC	Discrimination
ELPD-LOO	Expected Log (pointwise) Predictive Density using Leave-One-Out (LOO) cross-validation
EMHP	Environmental Monitoring for Health Protection
ENSO	El Niño Southern Oscillation
ECDC	European Centre for Disease Control
ETL	Expected Tail Loss
EW	Equally Weighted
EWS	Early Warning System
FoI	Force of Infection
GAM	Generalised Additive Model
gc/l	Gene copies per litre
GCV	Generalised Cross-Validation
GNN	Graph Neural Network
GP	Gaussian Process

Acronym	Description
GPE	Geospatial Population Estimate
HMC	Hamiltonian Monte Carlo
IBGE	Brazilian Institute of Geography and Statistics
ICEN	El Niño Coastal Index
ICU	Intensive Care Unit
IGP	Geophysical Institute of Peru
IID	Independent and Identically Distributed
INEI	National Institute of Statistics and Information of Peru
INLA	Integrated Nested Laplace Approximation
IQR	Interquartile Range
$\mathbb{K}$	Next-Generation Operator/Matrix
KL	Kullback-Leibler
LMIC	Low- and Middle-Income Country
LOO	Leave-One-Out
LOO-CV	Leave-One-Out-Cross-Validation
LSOA	Lower Layer Super Output Area
LSTM	Long Short-Term Memory
LTLA	Lower Tier Local Authority
MAE	Mean Absolute Error
MAUP	Modifiable Areal Unit Problem
MCMC	Markov Chain Monte Carlo
ML	Machine Learning
MSC	Miscalibration
NegBin	Negative Binomial
NGM	Next-Generation Matrix
NGO	Next-Generation Operator
NHS	National Health Service
NPI	Non-Pharmaceutical Intervention
ODE	Ordinary Differential Equation
ONI	Oceanic Niño Index
ONS	Office for National Statistics
PDE	Partial Differential Equation
PE	Permutation Entropy
PI	Prediction Interval
PIML	Physics-Informed Machine Learning
PINN	Physics-Informed Neural Network
PIT	Probability Integral Transform
PL	Pinball Loss
PMCMC	Particle Markov Chain Monte Carlo
R_0	Basic reproduction number
$R(t)$	Instantaneous reproduction number
r	Intrinsic/exponential growth rate
r_t	Instantaneous growth rate

Acronym	Description
REACT-1	Real-time Assessment of Community Transmission
REV	Relative Economic Value
ROC	Receiver Operating Characteristic
RMSE	Root Mean Square Error
RNA	Ribonucleic acid
RR	Relative Risk
RSI	Relative Strength Index
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
RT-qPCR	Reverse Transcriptase quantitative Polymerase Chain Reaction
RW	Random Walk
rWIS	Relative Weighted Interval Score
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SDE	Stochastic Differential Equation
SDM	Social Distancing Measure
SHAP	SHapley Additive exPlanations
SIMDA	Sistema de Monitoramento Diário de Agravos (Daily Disease Occurrence Monitoring System)
SIR	Susceptible-Infected-Removed
SMC	Sequential Monte Carlo
SPI-6	Standardized Precipitation Index (6-Month)
SSA	Stochastic Simulation Algorithm
SST	Sea Surface Temperature
STW	Sewage treatment works
TCN	Temporal Convolutional Network
TimeGPT	Time Generative Pre-trained Transformer
TLoD	Theoretical Limit of Detection
UK	United Kingdom
UKHSA	United Kingdom Health Security Agency
UNC	Uncertainty
US	United States of America
VOC	Variants of Concern
VOI	Value of Information
VUI	Variants Under Investigation
WAIC	Widely Applicable Information Criterion
WBE	Wastewater-Based Epidemiology
WHO	World Health Organisation
WIS	Weighted Interval Score
WWII	World War II
XAI	Explainable Artificial Intelligence

Chapter 1

Introduction

1.1 DPhil thesis background

Infectious diseases can be deadly and complex, but understanding and predicting their spread has never been more crucial. In doing so, we can quantitatively inform real-time decision-making and long-term public health policy, both of which have far-reaching consequences for epidemic control and for global human, animal, and environmental health[3,4,5,6,7].

The COVID-19 pandemic encapsulates the impacts of infectious diseases. Since the emergence of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus (the infectious agent that causes COVID-19) in late 2019, there have been implications for health, education, the environment, the economy, and many other aspects of society. There have been 7,096,187 deaths due to COVID-19 reported to the World Health Organisation (WHO) as of 5 June 2025, with many further health consequences for mental health, cancer, maternal and infant health, and beyond[8,9,10,11,12,13,14].

This global pandemic challenged scientists to support public health authorities in their efforts to combat the spread of the highly transmissible SARS-CoV-2 virus. The role of scientists was not limited to the development of effective medical therapeutics and vaccines, as infectious disease modellers were also placed at the forefront of policy design[15,16,17,18]. Analyses and models were routinely used to inform difficult, urgent, and uncertain decisions, including the design of surveillance and control strategies to monitor and reduce SARS-CoV-2 transmission and COVID-19 mortality[17,18,19]. The need for reliable and timely quantitative insights accelerated the evolution of modelling techniques (from semi-mechanistic models to reproduction number estimation techniques), and the expansion of the types and volume of data used (from government response indicators to mobility data)[15,20]. This allowed modellers to investigate many key questions held by public health authorities, including: How transmissible are the SARS-CoV-2 virus, its variants, and its sub-variants? How many people are currently infected with the virus? Is the epidemic shrinking or expanding? Which non-pharmaceutical interventions (NPIs) could most effectively reduce community transmission, healthcare pressures, and/or mortality?

The COVID-19 pandemic is a single, albeit likely immeasurably devastating, example of an infectious disease outbreak, and this recent memory has heightened awareness at a societal and governmental level of the uses, strengths, and limitations of infectious disease modelling[16,18]. Unfortunately, this pandemic also underlined the persistence of many inequalities, both within countries and globally, in transmission risk[21,22], control measures[23], funding[24], testing[25], health outcomes[26,27], and data availability[28]. These systemic inequalities are sadly not unique to the COVID-19 pandemic, as the spread of many infectious diseases has been amplified by colonisation, slavery, poverty, and war[29]. The global burden of infectious diseases, from malaria[30] and dengue[31] to HIV[32] and tuberculosis[33], has long been unequal and continues to be felt most heavily in low- and middle-income countries (LMICs)[34,35,36]. In such resource-limited settings, data collection is both economically and logistically challenging, which means that the data modellers rely on are scarce and highly imperfect, suffering from representativeness and timeliness issues, especially in the marginalised communities that diseases often

most disproportionately affect[6,35].

Similarly, the unfolding climate crisis, paired with increasing global connectivity and urbanisation, has increased the global distribution and incidence of climate-sensitive infectious diseases, whose impact is often greatest in tropical and sub-tropical LMICs. With these increases projected to continue in the future[37,38,39,40], it is necessary for scientific researchers to work collectively across branches of science and with local researchers and public health authorities to develop scalable methods that can provide reliable, timely, and useful information about the past, present, and likely future epidemic dynamics[3]. Equipped with such information, more equitable, targeted, and cost-effective public health policies can be designed, whether it be surveillance systems, control measures, or vaccination strategies.

Many other infectious pathogens (aside from SARS-CoV-2) have deadly and costly endemic impacts on society and/or have pandemic potential. In this thesis, we focus primarily on dengue, the world's most widespread and rapidly expanding mosquito-borne disease. Dengue has a familiar unequal global burden with the annual reported cases of over 14 million in 2024 primarily occurring in LMICs[41,42]. Scientists are therefore needed to support public health authorities in their fight to address the major and growing challenges to global health posed by the expanding distribution and increasing incidence of dengue.

Researchers can reflect on the COVID-19 pandemic, adopt lessons learned, extend scientific advances made, and shed light on poorly understood, evolving dynamics of other infectious diseases. For example, in the next epidemic of an unknown, re-emerging, or endemic disease, there will be known unknowns and unknown unknowns. However, are there generalisable, scalable methods and procedures that researchers can transfer, adapt, or implement directly? Similarly, can modellers identify how best to support the decisions of public health authorities with limited resources as they work tirelessly to reduce disease burden? These are questions that should be explored now to develop methods and insights that improve our preparedness and public health responses for future outbreaks, before epidemic control becomes infeasible.

1.2 DPhil thesis aims and objectives

The aim of this DPhil thesis is to connect methodological rigour with insights for public health policy. The research objectives (detailed for individual chapters in Table 1.1) are to present a suite of modelling theory, analyses, and tools that introduce new methodology, unify existing techniques, and ultimately extend beyond existing modelling approaches. To this end, in this thesis, focusing mainly on dengue in resource-limited settings, we will blend heterogeneous data sources, branches of science, and modelling disciplines to provide problem-led methods and analyses to better understand epidemic dynamics and drivers.

Working in this way, our aim (Figure 1.1) is to enable a greater volume and diversity of robust outputs that i) quantify the relative importance of *past* epidemic drivers, ii) monitor *present* epidemic dynamics, and iii) estimate likely *future* epidemic trajectories. To maximise the utility and strength of scientific evidence, we will adopt an integrated approach to modelling, as we will use our insights from models of the past, present, and future dynamics to inform model development and evidence synthesis across these different tasks.

While we are in a time of considerable and inequitable public health challenges (Section 1.1), we are also in an exciting time i) for research methods across mathematics, statistics, and artificial intelligence (AI), ii) for data with increasing availability, variety, and coverage, and iii) for scientific advancements more generally with new vaccine technology, control options, and treatment methods. However, achieving the objectives of this thesis, through new methods and analyses, depends on considerable data collection efforts across society and the shared knowledge of individuals and authorities in often resource-limited settings, and we hope that the aspired public health utility of our research reflects such efforts. Indeed, our research objectives are grounded in practical and public health-focused objectives. For example, we echo the objectives of the WHO Global Arbovirus Initiative throughout the thesis: to develop new approaches, proactively monitor and reduce epidemic risk, enhance vector control strategies and epidemic preparedness, and develop local coalitions[43].

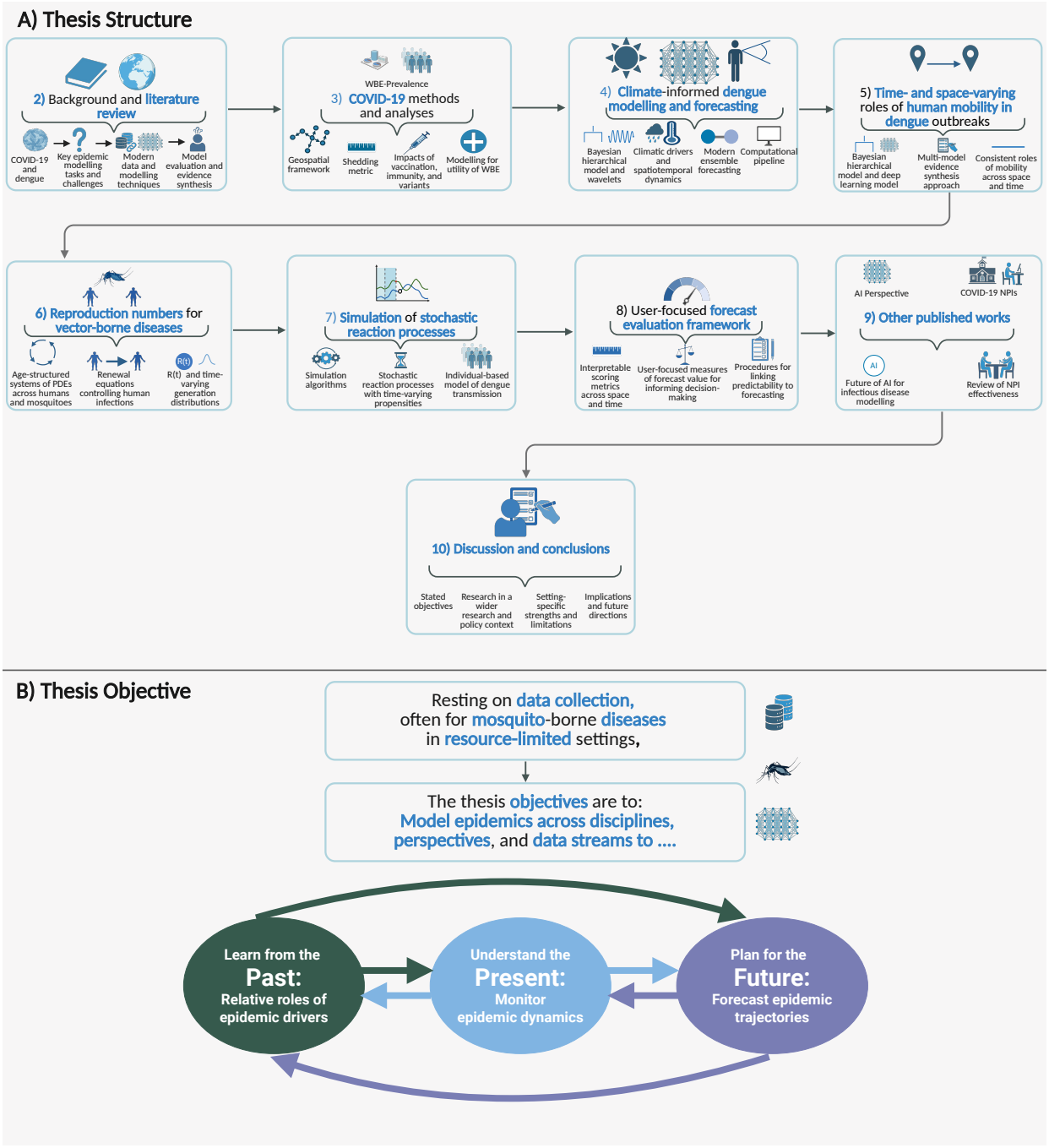


Figure 1.1: **Graphical abstract and thesis objectives**. In **A)**, we visualise the flow of chapters in this integrated thesis and the high-level contributions. In Table 1.1, we provide further details of both the research questions and contributions of individual chapters. **B)** In this thesis, we aimed to develop models to i) learn from the past, ii) understand the present, and iii) plan for the future. We will argue that the insights from each modelling endeavour can inform one other, and alongside our multiple models and diverse data streams, can ultimately provide a greater variety of robust information for public health authorities. In Chapter 10, we will discuss the contributions of the individual chapters towards the thesis objectives in a wider research and public health context (see Figure 10.1).

Adopting this consistent focus on public health priorities, we aim for actionability and reliability in our theoretical frameworks and analyses such that they can be used to reliably inform real-time decision-making and long-term public health policy, for current and future infectious disease threats.

1.3 DPhil thesis structure

First, I would like to acknowledge here the contributions of those who participated in the collection, collation, and sharing of data, without whom much of this research would not be possible. We summarise the structure of this integrated thesis in Table 1.1, alongside our research questions for each chapter and our contributions to answering these questions.

There are seven manuscripts included in full in this thesis, alongside abstracts for two other published works. All full manuscripts are first-authored by me and, aside from minor formatting changes, are presented in their published/submitted format for consistency. We now provide further details of the main chapters:

- **Chapter 3: The utility of wastewater surveillance for monitoring SARS-CoV-2 prevalence** This chapter presents a manuscript, published in *PNAS Nexus*[44], that explores how best to use wastewater surveillance for monitoring the prevalence of SARS-CoV-2 in communities across a nation.
- **Chapter 4: Unifying techniques for climate-informed dengue modelling and forecasting** This chapter presents two manuscripts for modelling dengue epidemics across space and time in northern Peru. The first manuscript, published in *PLOS Neglected Diseases*[45], introduces methods to probabilistically quantify multiple climatic influences on dengue incidence, which then informs methods to probabilistically forecast of future dengue incidence. In a follow-up manuscript, currently under review at *Royal Society Open Science*[46], we extended these methods to a multi-model workflow for modelling past dynamics and drivers and forecasting future dengue incidence at finer spatial resolutions.
- **Chapter 5: The time- and space-varying roles of human mobility in shaping urban dengue outbreaks** This chapter presents a manuscript, currently under review at *PNAS Nexus*[47], where we developed a multi-model evidence synthesis approach to quantify the relative importance of human movement in urban dengue outbreaks across space and time.
- **Chapter 6: Reproduction numbers for vector-borne diseases** This chapter lays theoretical foundations for estimating reproduction numbers for vector-borne diseases. The manuscript, currently under editorial review at *Methods in Ecology and Evolution* following minor revisions[48]. We introduced an age-structured modelling framework of partial differential equations (PDEs) across mosquito and human populations, which allowed us to derive renewal equations that describe how past human infections generate new human infections.
- **Chapter 7: Exact simulation of stochastic reaction processes with external variability** In a related manuscript to be submitted to *The Journal of the Royal Society Interface*[49], motivated by the need to simulate our system of PDEs, this chapter presents extensions of the Gillespie algorithms[50,51] for simulating stochastic reaction processes with time-varying propensities between events.
- **Chapter 8: From metric to action – An evaluation framework to translate infectious disease forecasts into policy decisions** Blending ideas from statistics, decision theory, and weather forecasting, this chapter introduces a new evaluation framework for forecasters and forecast evaluators that shifts the perspective of forecast evaluation to the decision-maker. The manuscript will be submitted to *The International Journal of Forecasting*[52].
- **Chapter 9: Other published works** This chapter presents two published manuscripts. In the first manuscript, published in *Nature*[53], we explored the future of Artificial Intelligence (AI) for epidemiological modelling. The second manuscript is a systematic review commissioned by *The Royal Society*[54] as part of the theme issue "The effectiveness of non-pharmaceutical interventions on the COVID-19 pandemic: the evidence". We reviewed the effectiveness of various social distancing measures and lockdowns during the COVID-19 pandemic.

Table 1.1: **Research questions and contributions by chapter:** We outline here the research questions for each chapter of this integrated thesis, and summarise their individual contributions to satisfying these research gaps. Each of these research questions and contributions are grounded in modelling theory, yet the modelling objectives remain public-health-focused. Figure 1.1 provides a visualisation of how each of these contributions fit into the wider thesis objective of developing models to learn from the past, monitor the present, and planning for the future. In Chapter 10, we will discuss further the contributions of these individual chapters towards the thesis objectives (see also Figure 10.1) in a wider research and public health context.

Chapter	Research questions	Chapter contribution
Chapter 3 (Wastewater surveillance and SARS-CoV-2 prevalence)	What is the relationship between SARS-CoV-2 wastewater concentrations and community prevalence? How can wastewater-based epidemiology (WBE) best be used for monitoring community prevalence in high spatial resolution?	A new faecal shedding metric that unveils the impacts of vaccines, immunity, and variants on population-level shedding A geospatial framework and predictive model for prevalence as a function of wastewater concentrations A likely complementary role for WBE in monitoring community prevalence
Chapter 4 (Climate-informed modelling and forecasting of dengue)	How to quantify spatiotemporal patterns and climatic drivers of dengue epidemics in Peru? What is the structure, timing, and intensity of delayed climatic effects on dengue epidemics at different resolutions? How best to reliably forecast future short-term dengue epidemic trajectories and evaluate these forecasts? How to answer these questions with multiple models?	Unified wavelet and Bayesian hierarchical methods in a workflow for retrospective analyses Multi-model quantification of spatiotemporal epidemic dynamics and of climatic influences on dengue incidence in Peru Climate-informed forecasting models and ensembles with outbreak detection indicators Generalisable computational pipeline for simultaneous, real-time modelling of past and likely future dengue epidemic dynamics
Chapter 5 (Relative roles of human mobility in urban dengue outbreaks)	For dengue outbreaks in an urban setting (Fortaleza, Brazil): How does the relative importance of human mobility vary over time and across space? How does the relative influence of different neighbourhoods vary over time? How can we answer these questions with multiple models?	Multi-model evidence synthesis framework Consistent relative importance of human movement for establishment, expansion, and persistence of dengue outbreaks Consistent spatial profile of neighbourhoods which drive dengue transmission
Chapter 6 (Reproduction numbers for vector-borne diseases)	How to measure time-varying transmissibility of vector-borne diseases? What is the meaning of a time-varying generation time distribution? How to exactly simulate stochastic reaction processes with time-dependent rates?	Age-structured system of PDEs that enable derivation of new renewal equations for diseases with multiple stages in the transmission cycle Mechanistic model and simulation theory for future inferential frameworks for $R(t)$ Meaning from mechanistic formulae for time-varying generation time distributions and $R(t)$
Chapter 7 (Simulation of stochastic reaction processes)	How to exactly simulate stochastic reaction processes that experience external variability and therefore, time-dependent propensities between events?	Two algorithms, building on seminal Gillespie algorithms [50,51], which allow for time dependencies in propensities between events Simulations for systems of chemical reactions and individual-based models of dengue transmission
Chapter 8 (An evaluation framework to translate infectious disease forecasts into policy decisions)	How to measure value of forecast for informing decision-makers? How does predictability of epidemics shape the value of forecasting? How can decision-makers take appropriate action from a forecast?	New measures of forecast value and predictability analyses that convert metrics to actionable and reliable outputs for decision-making Evaluation workflow for forecasters and forecast evaluators that prioritises decision-makers, statistical rigour, and public health utility
Chapter 9 (Other published works)	Artificial Intelligence (AI): How best to use AI for epidemiological modelling? COVID-19: How effective were social distancing measures and lockdowns at reducing community transmission?	Uses of AI for modelling climate-sensitive infectious diseases, and more generally statistical inference, forecasting and prediction, and the integration of modern/novel data sources. A systematic review where we found that measures varied in effectiveness and the most stringent measures, often implemented concurrently, were most effective.
Chapter 10 (Discussion and conclusions)	How have the chapter contributions allowed us to achieve our stated objectives? How do these results relate and contribute to the wider literature base? How can we build on our theory and analyses?	Learning from the past: The wider context of our multi-model inferences about mobility and climatic effects on dengue epidemics, and future directions for mechanistic models, joint inference, and operationalisation. Monitoring the present: The theoretical and practical consequences of our methods focused on $R(t)$ estimation for vector-borne diseases, simulation, and WBE, alongside future public-health-focused extensions. Planning for the future: The contributions of our models and evaluation framework towards more reliable and actionable epidemic forecasting, alongside details of future directions, considerations, and operationalisation.

Chapter 2

Background and literature review

This thesis is largely statistical and mathematical by construction. However, before exploring fundamental questions of *why* and *how* we model infectious diseases, we will also briefly describe the epidemiology of the infectious diseases studied in this thesis. The associated biological, environmental, and human factors involved in the complex processes of infection will influence our proposed methods for modelling epidemic dynamics and synthesising evidence for scientific and policy understanding.

This chapter is therefore structured by: i) the epidemiology for the studied infectious diseases, ii) the motivations for modelling infectious diseases and epidemics, iii) key epidemiological modelling tasks, iv) challenges for these modelling tasks, v) complementary data streams to address the challenges, vi) techniques for modelling epidemics, and vii) the final steps of model evaluation and evidence synthesis.

2.1 Epidemiology of COVID-19 and dengue

2.1.1 COVID-19

SARS-CoV-2 is an enveloped, positive single-stranded RNA virus. Containing nearly 30,000 letters of RNA, it is an airborne contagious virus that is primarily transmitted by respiratory droplets exhaled by an infected individual when breathing, talking, or coughing[55]. Less common transmission routes include indirect transmission via contaminated objects. SARS-CoV-2 can be spread by presymptomatic, symptomatic, and asymptomatic individuals, and the incubation period (time from pathogen exposure to symptom onset) across different strains is estimated to be 6.6 days (95% Confidence Interval (CI): 6.3, 6.9)[56]. The clinical and epidemiological characteristics of the COVID-19 disease include a wide variety of symptoms, ranging from mild respiratory distress to severe pneumonia. COVID-19 induces greater risk of mortality in older and more vulnerable populations and is the most prominent recent example of a pandemic-inducing infectious disease with substantial, long-lasting global consequences[57].

Faced with exponential growth in the number of new infections in the early stages of the COVID-19 pandemic, alongside no suitable medical therapeutics or effective vaccine, governments and public health authorities designed and implemented measures to control pathogen spread. Public health measures, ranging from school closures to nationwide lockdowns, varied across space and time in their stringency and duration, and it was often through the lens of epidemiological models that interventions were designed, their timing informed, and their effects assessed[17,18,58,59,60,61].

Consistent with any pandemic caused by an infectious pathogen, these NPIs were designed and implemented to reduce pathogen transmission, exceptional healthcare pressures, and disease-related mortality. Even for past outbreaks, understanding the "success" of public health measures (often by estimating reductions in such epidemiological quantities) is

important for guiding current and future public health policy. Their success, while potentially difficult to quantify causally in observational settings of epidemics, is determined by a coalescence of human (e.g., public compliance), biological (e.g., variants of SARS-CoV-2), and environmental (e.g., locations of infected individuals) factors[4,5,62].

2.1.2 Dengue

Unlike COVID-19 which is transmitted from human to human, dengue is a mosquito-borne disease caused by one of four (known) antigenically related but distinct serotypes of dengue virus (DENV 1–4)[64]. DENV has a long history of transmission, as it was first isolated in the 1940s[65,66]. Known circulation stretches over 400 years and the first record of a clinically compatible case stems from a Chinese medical encyclopaedia in 992 AD[63,65]. DENV is a positive-stranded RNA virus and a member of the Flaviviridae family which also includes Zika, West Nile Virus, and yellow fever. The primary route of DENV transmission to humans is via infected *Aedes* mosquitoes, which are daytime biting mosquitoes that thrive in urban and peri-urban areas[67,68,69]. Adult female *Ae. aegypti* are the principal vectors, while *Ae. albopictus* are generally secondary vectors[70]. Rarer routes of transmission include blood transfusion, perinatal transmission, and organ transplants[71].

Symptoms of dengue are wide-ranging, including fever, nausea, joint and muscle pain, and headache. All four DENV serotypes can cause human infection and exposure to a single serotype offers lifelong exposure to that serotype and short-lived (< 2 years) cross-protection to other serotypes. 40–80% (mean estimate = 54% across 50 studies)[72,73] of all DENV infections are asymptomatic or result in mild illness, meaning that it can potentially be self-treated. This makes infections difficult to detect by routine surveillance. A small fraction of individuals develop severe dengue with potentially life-threatening consequences (e.g., a case fatality rate in 2023 of 0.11%)[67,74]. Various viral and host theories have been proposed for the development of severe disease, including antibody-dependent enhancement (ADE), where an individual's antibodies from a previous infection caused by another DENV serotype enhances the entry of DENV in a secondary infection into the host's cells[75].

Endemic in more than 100 nations, dengue poses an increasingly global public health threat, with an estimated 5.66 billion people now living in areas suitable for dengue transmission[69]. Dengue induces both endemic and epidemic transmission patterns in urban and rural settings and has considerable societal and economic impacts[71], with macroeconomic impacts (e.g., absenteeism in the workforce) estimated to far outweigh medical costs[7]. DENV is now the world's most common and rapidly spreading arbovirus (i.e., arthropod-borne virus), and a record number of over 14 million cases were reported in 2024, with the most pronounced increases in the Americas (especially Brazil, Colombia, Mexico, and Peru)[31,41,42]. While predominantly found in tropical and subtropical regions, with globalisation, climate change, and urbanisation, the global distribution and incidence of dengue disease cases are expanding and rising respectively, both in endemic and new areas (Figure 2.1 D). These factors create increasingly suitable conditions around the world for viral transmission and mosquito development, reproduction, and survival[72,76,77,78,79]. In fact, the first historical expansions and increases identified were attributed to the post-World War II (WWII) trends of increased globalisation (aided by modern air transport), population growth, and unplanned urbanisation[77]. This increasingly brought more humans into close contact with both mosquitoes and DENV.

Dengue is a climate-sensitive infectious disease, as the mosquito (female *Aedes* mosquitoes) and pathogen (DENV) are climate-sensitive, meaning that climatic and weather variability influences the abundance, distribution, and activity of mosquitoes and the generally seasonal transmission of DENV[70,80,81,82,83]. For example, the reproduction and development of the mosquito and pathogen are temperature-dependent, while mosquitoes require sufficient precipitation to lay eggs (Figure 2.1 A). These sensitivities suggest how climate-informed modelling could be used to understand, monitor and forecast past, current, and future trends in dengue epidemics. However, like any climate-sensitive infectious disease, the effects of climatic conditions on DENV transmission are modulated by biological and human influences. For example,

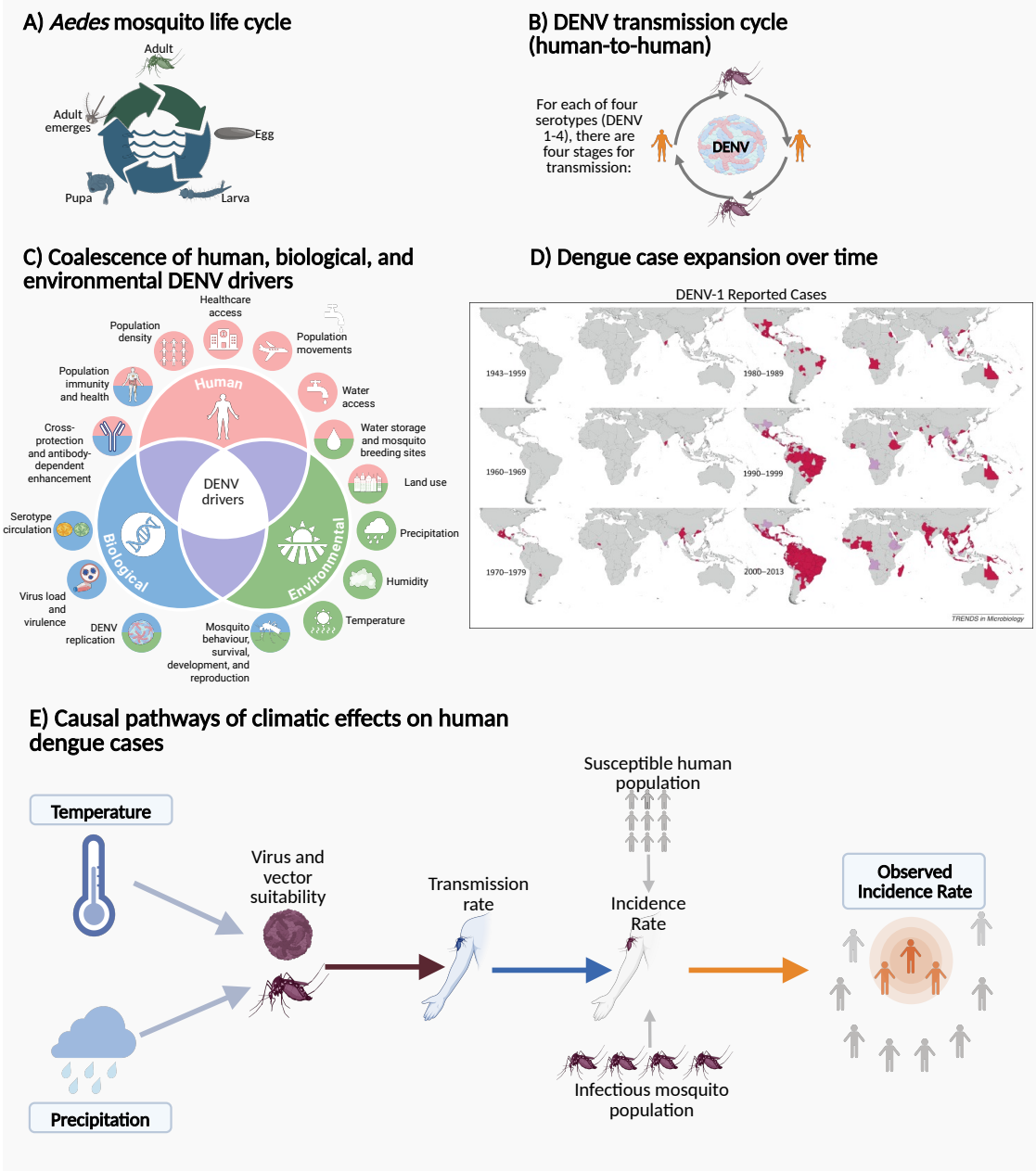


Figure 2.1: **Infographic for dengue**. Across this thesis, we often provide theory and analyses for dengue, the mosquito-borne disease caused by infection with dengue virus (DENV). This visualisation captures many important concepts in DENV transmission necessary for understanding transmission, modelling transmission, and synthesising evidence. **A):** We present a simplified version of the *Aedes* mosquito life cycle from egg to larva to pupa to emergence of the adult mosquito. **B):** For any of DENV 1–4 to be transmitted from human to human, a susceptible mosquito must bite an infectious human and then become infectious before biting another susceptible human. **C):** DENV transmission risk is influenced by many biological, human, and environmental factors. Many of the visualised influences could be considered at the intersection of these three broad categories. **D):** Reported dengue cases globally have both risen in number and expanded in distribution over the past 80 years. We provide a copy of Figure 1 from Messina et al. [63], where darker colour indicates areas where DENV-1 cases were reported in the given decade and lighter colour indicates areas where DENV-1 cases had previously been reported but not in the given decade. **E):** A simple visualisation of a common causal setting faced in studies of this thesis, where items shaded in blue are often the observed variables, and we do not have data or even estimates for many of the unobserved variables. For simplicity, we have not included other crucial drivers of vector-borne transmission such as host preferences. Here, we would be estimating the *direct* effect of temperature and/or precipitation on the observed dengue incidence rate, though some of the modulators could be indirectly modelled (e.g., reporting rates) by using spatial and/or temporal random effects.

water storage practices during drought periods can exacerbate the risk of DENV infection due to the presence of more stagnant water (e.g., in tires or artificial containers) that is favoured by *Aedes* mosquitoes for breeding sites [78,84]. Similarly, human movement can promote DENV importations, serotype introductions and reintroductions, and continued spread across geographies [85,86]. Despite such human and biological confounders, and interaction effects, models of dengue sometimes rely on estimates of temperature dependencies for mosquito and viral traits [80,82,87,88] derived from highly controlled, laboratory studies. These underpin many model-based estimates of climatic dependencies in epidemics and the impacts of climate change [80,82,87,88,89]. However, faced with an increasingly global public health threat caused by severe dengue and side-effects, it important that scientists provide decision-makers with a more accurate scientific understanding of environmental, human, and biological influences on complex dengue epidemic dynamics. Without such understanding of how mosquito and viral traits respond in settings beyond the laboratory, designing effective control strategies from a model's outputs is less feasible.

In terms of policy options, in the absence of effective, dengue-specific vaccines or antiviral treatments for humans (see Chapter 10.2), dengue interventions have historically been indirect and focused on vector control. These include indoor and outdoor spraying of insecticides (targeted at adult mosquitoes), removal of oviposition sites, and larvicide usage (targeted at egg, pupal, and larval stages), all of which have had little success for dengue epidemic control [71,90,91,92,93]. Historically, dichlorodiphenyltrichloroethane (DDT) was a vector control strategy developed during WWII. Although it successfully reduced *Ae. aegypti* populations, due to environmental and health impacts and the presence of DDT-resistant mosquitoes, DDT is now prohibited in most countries. More recent novel control strategies include Wolbachia-based incompatible infected mosquitoes and sterile mosquitoes. These have indicated early promise for effective vector control [71,94]. Integrated vector management is the recommended strategy of the WHO, and this strategy promotes the combination of multiple environmental, biological and chemical control methods. Their ultimate success again depends on both human factors (e.g., funding, implementation, and population responses) and biological factors (e.g., *Aedes* resistance and virus circulation) [95,96,97].

2.2 Fundamental question: *Why* model infectious disease outbreaks?

The intricate yet fundamental biological, mathematical, and statistical details of a model often matter so little to so many. Stepping back from a modeller's perspective, we should first question the utility and purpose of any modelling exercise. Epidemiological modelling can only generate abstractions of the unknown, complex reality, and so, which questions can (and should [20]) we answer?

Modelling[†] is not purely a theoretical exercise, as the need to understand *why* and *how* infectious diseases spread is driven by the need to inform public health policy and advance scientific understanding [62,100,101]. A well-validated, model-based understanding of epidemic dynamics can be used, both before and during outbreaks, to reliably inform long-term public health planning (e.g., Intensive Care Unit (ICU) capacity planning) and real-time decision-making (e.g., proactive and reactive interventions) [3,4,5]. Given the intensive data collection efforts of those in affected areas, the societal impacts of public health policy, and the often-innumerable impacts of an infectious disease itself, it is clear that epidemiological models must be based on rigorous methodology.

Practically, public health authorities often use the outputs of models when designing policies that require an understanding of many different unknowns [17,102,103], including recent spatiotemporal epidemic patterns, the real-time transmissibility of a pathogen, the epidemic drivers (environmental, biological and human influences), and the likely current (nowcasting)

[†] In this thesis, we use the term *model* very generally, reflecting a perspective from mathematical biology that a model can be a quantitative description or a set of hypotheses about data (observed or simulated) or, more generally, about a system [98]. Heesterbeek et al. [99] would label our model definition as a *quantitative approach*.

and future short- and long-term epidemic trajectories. For example, imagine a new zoonotic pathogen or variant has been detected by genomic or wastewater surveillance. How transmissible is the emerging pathogen or variant? Given the resources allocated for surveillance, how many infections occurred before the first detection? Will the pathogen or variant cause a widespread epidemic? If so, how has it spread and why? Where should diagnostic tests be targeted and at which demographics? What are the likely future epidemic trajectories and how can we target different epidemic control measures and surveillance techniques most equitably and effectively? These are some of the many questions that robust epidemiological models, together with data and ideas from many branches of science, can help answer (Figures 2.2 and 2.3).

We conclude this interlude on the potential of epidemiological models by emphasising cautionary notes that i) a model is an incomplete abstraction of reality that is constrained by our imperfect knowledge of the world, and so ii) a model and its outputs must be rigorously designed in a timely manner to inform public health policy, tested for reliability and user value, and then interpreted and communicated correspondingly to prevent model misuse[20,100,103,104,105]. As Pablo Picasso once said (apparently[106]): "Art is a lie that helps us to discover the truth."

2.3 Epidemiological data and derived quantities

During an epidemic, we never directly observe *true* epidemiological quantities (e.g., infections), instead relying on observational data, collected by local authorities, in the form of fundamental indicators such as reported disease cases, deaths, hospitalisations, or estimated prevalence[15,107,108,109]. While often biased[110,111], incomplete[15,107,112], delayed[113,114], and subject to time-and space-varying quality problems[115] (see Section 2.4), these routinely collected aggregate measurements are stepping stones to providing public health authorities with a model-based understanding of past, current, and future epidemic dynamics. See Table 2.1 for some examples of key epidemiological quantities when modelling dengue epidemics.

2.3.1 Informing early outbreak response

On their own, the utility of epidemiological surveillance data for informing public health policy can be limited, providing no direct information about the strength or speed of an epidemic (see below) and/or misrepresenting true numbers (e.g., infections if asymptomatic transmission is poorly understood or captured)[4,15]. For example, the prevalent disease cases at a time point can be generated by many different epidemic trajectories and different transmission patterns, which means that authorities do not provide necessary information about *how fast* (i.e., transmissibility) or *why* a pathogen is spreading. To address such issues, modellers often use data, alongside a mathematical/statistical abstraction (a model) of epidemic dynamics, to derive related quantities from the fundamental indicators.

Understanding the transmission potential of a pathogen is crucial. When a new pathogen emerges, public health authorities need to know the invasibility and transmissibility of the pathogen, and, if established, the likely size of the outbreak and future epidemic trajectories[15,19]. The mathematical quantity R_0 is the basic reproduction number – the average number of individuals infected by an average primary case in a fully susceptible population over the infectious lifetime of the primary case[116,117]. R_0 is dependent on both the pathogen (and/or its variants) and the given population and is only defined at the start of an outbreak[118,119]. Mathematically, R_0 depends on the model and its assumptions (see below)[120,121]. This quantity may appear abstract, yet it has many uses. For directly transmitted infectious diseases in homogeneously mixing populations, R_0 can be used for i) the estimated threshold for pathogen establishment (if $R_0 > 1$, a pathogen can likely invade), ii) the estimated threshold for vaccine-induced herd immunity ($1 - 1/R_0$), and iii) the estimated final epidemic size ($1 - \exp(-R_0)$)[122,123].

Mean early pandemic estimates of R_0 for the ancestral SARS-CoV-2 in China were 3.28[124], but consistent with the dependencies on the population and the structure of the model, the estimates varied between populations (e.g., mean

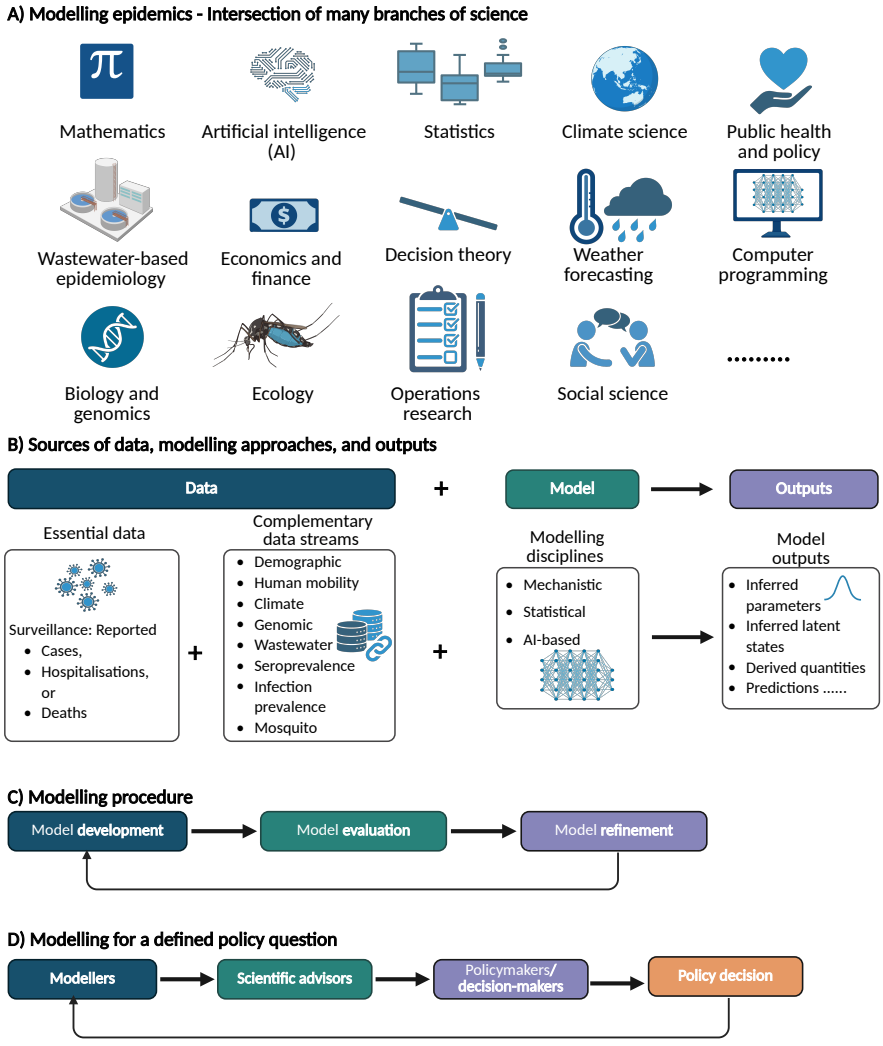


Figure 2.2: **Modelling for scientific understanding and public health policy**. Throughout this thesis, the theory and analyses of chapters follow similar yet distinct themes. Each chapter is motivated by the research objective of developing methods and analyses that can inform public health policy, often in resource-limited settings. **A):** This frequently relies on epidemiological modelling at the intersection of many branches of science, across both the methods and data, and these are introduced at a high level throughout Chapter 2. **B):** As discussed in Sections 2.3 – 2.5, there are fundamental indicators of epidemic dynamics and complementary data from heterogeneous data streams, all of which can be integrated as imperfect but useful inputs to different types of models to achieve the targeted model outputs for different policy/research questions. **C):** The modelling procedure, across various approaches, creates abstractions of the epidemiological reality in different ways, yet at a high-level, it is an iterative process involving development, evaluation, and refinement. This is the common setting of the current thesis **D):** In real-world settings beyond this thesis, modelling for a defined policy question ideally involves communication of evidence and scientific understanding among stakeholders before a policy decision is taken. Modellers can then incorporate this policy into their models and assess retrospectively the impacts of such policy decisions on epidemic dynamics and wider society. This high-level overview is complemented by Sections 2.2 – 2.7, and summaries of data streams in Table 2.2, of modelling classes in Figure 2.3, and evaluation in Figure 2.4.

$R_0 = 4.22$ in the European Union at the early stages [125]) and variants (e.g., mean $R_0 = 5.08$ for the Delta variant in the United Kingdom (UK) and China [126]). For DENV, Liu et al. [127] found a mean average $R_0 = 2.55$ across 54 studies in tropical, sub-tropical, and temperate regions, although published estimates ranged considerably from 0.16 to 65.

R_0 is a unitless quantity that can be calculated using an epidemiological model and/or the spectral radius $\rho(K)$ of the next-generation operator (NGO) K [116,128]. The NGO describes the number of infections occurring in successive generations

Table 2.1: **Key epidemiological parameters inferred using dengue models.** Focusing on dengue, the primary disease studied in this thesis, we list several types of model parameters (left column), complemented by examples of each (right column). Often relying on more data streams beyond epidemiological surveillance data, these model parameters can be inferred using one or more of statistical, mechanistic, and/or AI-based models.

Parameter type	**Examples**
Basic reproduction number	Reproduction numbers for a system of dengue transmission and for human-to-human transmission, R_0
Time-varying reproduction number	Instantaneous or case reproduction numbers $R(t)$ for human-to-human transmission
Epidemic growth rates	Intrinsic (r) and instantaneous growth rates (r_t) in human infections
Hidden states	Susceptible, exposed, infectious, and recovered states across human and mosquito populations
Fixed/random effects	Environmental, biological, human, spatial, and temporal effects on human dengue cases or deaths
Transmission model parameters	Force of infection, recovery rates, development, survival, and reproduction rates across human and mosquito populations and/or DENV
Observed process parameters	Reporting rate, clinical detectability, delay, and censoring parameters
Intervention parameters	Efficacy of vaccination, vector control, and/or public information campaigns

of transmission and works by structuring populations into different states/ categories of infected individuals. For a discrete, finite number of states, the NGO becomes the next-generation matrix (NGM) $\mathbb{K}$ [116,128], and R_0 is simply the largest (in magnitude) eigenvalue of the NGM:

$$R_0 := \rho(\mathbb{K}) = \lim_{n\to\infty} (\mid \mathbb{K}^n \mid)^{\frac{1}{n}}, \tag{2.3.1}$$

where n denotes the generation number. This means that R_0 defines a population multiplication factor on a generation basis. Paradoxically, despite being applicable only for early outbreak dynamics, R_0 describes the asymptotic average next generation size. This tension, created by using long-term transmission dynamics for an early stage indicator, is an important yet often overlooked disclaimer with implications for conclusions about controllability (e.g., R_0 above/below 1) [129].

In this thesis, when estimating any reproduction number, we must clarify the setting-specific definition of new infections, possibly including both hosts (typically humans) and vectors or just hosts. We can define either a *human-to-human* basic reproduction number or a basic reproduction number for a *system* of infections across humans and vectors [119]. Similarly, care is required as R_0 is an expectation across one or more populations and can obscure individual-level heterogeneities in infectiousness [130] or in the propensity to acquire infection [104]. Nevertheless, timely and reliable estimates of a human-to-human R_0 can practically inform epidemic control strategies, from early outbreak responses to resource planning for later stages of an epidemic [4,131]. A key lesson from the COVID-19 pandemic and other epidemics was that public health actions must be taken rapidly in the early stages of exponential growth as prompt action, even by a few days, can be vital for averting new infections [132]. This highlights the potential for timely and reliable epidemiological models' estimates that inform public health policy.

For mosquito-borne diseases, various relative R_0 indicators have also been proposed as functions of temperature T [82,87,88]. Generally, $R_0(T)$ indicators use temperature-dependent mosquito (e.g., biting, egg-laying, and death rates) and viral (e.g., DENV infection probability) traits. These traits have been derived under different models for thermal responses (e.g., Brière/quadratic functions) which are fitted to controlled laboratory data. These $R_0(T)$ metrics differ from the above, as they are relative in the sense that values are scaled to lie between 0 and 1, where higher values indicate when transmission should be possible [82,83].

The intrinsic or exponential growth rate r is another time-variant quantity (similar to R_0). r describes the rate of

epidemic growth in the early stages of an outbreak when the entire population is susceptible. It is defined as the per capita rate of change in the number of new cases/infections, and can be estimated using a compartmental, growth, or statistical model alongside epidemiological surveillance data [117,133,134]. It is an asymptotic measure in the sense that it is the long-term rate of growth of the system.

While r measures epidemic *speed* (how fast the epidemic is growing) and R_0 measures epidemic *strength* (how many infections are generated), they are related as the $r = 0$ threshold is equivalent to $R_0 = 1$ under constant conditions [135]. However, R_0 and r can only measure controllability when interventions are implemented without delay and do not change the time between infections, and the controllability properties of R_0 and r can be impacted by individual-level variations (e.g., superspreaders) [129,130].

2.3.2 Measuring time-varying epidemic properties

Both r and R_0 help to quantify early outbreak dynamics, but do not account for time-varying transmissibility, short-term dynamics, changes in human behaviour, and depletion of the susceptible population if the epidemic has become established [119,129,136]. Epidemics are fast-evolving, and this leads to time-varying or effective reproduction numbers ($R(t)$) and instantaneous epidemic growth rates (r_t). There are various time-varying reproduction numbers [137]. The *case* reproduction number is the average number of secondary cases that is generated by a primary case infected at time t over the individual's entire lifetime of infection (i.e., after time t). The *instantaneous* reproduction number at time t measures the expected/predicted number of infections that would be generated by an individual (usually human) infected at time t throughout their infected lifetime if conditions affecting human transmission at time t were to remain the same [138,139,140]. As the case reproduction number relies on data or assumptions about future transmissibility after the time the quantity is estimated, it is a leading estimator of the instantaneous reproduction number.

In this thesis, we focus on the human-to-human instantaneous reproduction number (henceforth denoted $R(t)$) as we aim for a measure of instantaneous transmissibility that can be easily implemented in real time to inform epidemic control strategies. This differs from the case reproduction number which changes in a delayed, smooth manner [140]. Similar to R_0, $R(t)$ is a unitless quantity with identical threshold properties. At a basic level, control interventions can be designed to reduce the value of $R(t)$ below the threshold of 1, which suggests an epidemic under control and contracting. $R(t)$ entered widespread public conscience throughout the COVID-19 pandemic, as i) its threshold property allows modellers and public health authorities to track the evolution (past and current dynamics) of an epidemic, estimate the effectiveness of past interventions, and ii) its unitless nature allows comparison of transmission across different spatial scales, data streams, and epidemic phases [118,134,136].

However, $R(t)$ does not preserve/encode any temporal information, and therefore cannot measure the epidemic speed, unlike the instantaneous epidemic growth rate r_t. Empirically, r_t can be estimated from case data and is then defined as the rate of change of the log-transformed case incidence at a particular time point [134,136]. Estimates of r_t can also be derived for infections from transmission models [134]. In either derivation, r_t tracks epidemic speed over calendar times t, and measures both the magnitude *and* the timescale of spread. Primary uses of r_t include comparing the rate of spread of invading and existing variants and estimating doubling times (the time it takes for the number of infected individuals to double).

Both $R(t)$ and r_t can be useful, interpretable measures of time-varying epidemic dynamics, yet their utility is conditional on having some knowledge of infection prevalence [132]. For example, the utility of estimates for $R(t)$ at low prevalence levels can be limited by statistical variability, and values of $R(t)$ have different implications at different prevalence levels. Therefore, both $R(t)$ and r_t must be estimated and interpreted with caution. Estimates of $R(t)$ are interpreted for the *average* individual and depend on the data and model used including assumptions about the model structure, smoothing windows, and other parametric assumptions. [134,141,142,143,144]. Estimating r_t from case data is also not trivial, depending again on smoothing assumptions. r_t also does not provide the same public-health-focused implication for control measures

needed to prevent epidemic growth [134].

Controlling the relationship between $R(t)$ and r_t, and between R_0 and r, is the generation time distribution that acts as an infectiousness profile, describing the timescale of individual-level transmission [138]. It is the distribution of times between primary and secondary human infections and accounts for varying times between successive infections throughout the population. Assuming a well-specified, fixed (over calendar time) generation time distribution, the $R_0 = 1$ threshold corresponds with the $r = 0$ threshold [145].

In the absence of external surveillance data, for both statistical (changes in $R(t)$ and the generation time distribution are rarely identifiable) and practical (generation times are notoriously difficult to measure [62,146]) reasons, the generation time distribution is often assumed to be fixed over calendar time [145]. Mathematically, this leads to the probability density for the intrinsic generation time distribution $g(\tau)$ which describes the relative contributions to secondary infections by individuals infected τ time units ago [145,147]:

$$g(\tau) := \frac{K(\tau)}{R_0} \tag{2.3.2}$$

where $K(\tau)$ is the intrinsic infectiousness kernel (the rate at which secondary human infections are generated by individuals who were infected τ time units ago), and $R_0 = \int_0^\infty K(\tau)d\tau$ is the basic reproduction number. Then, Wallinga and Lipsitch [133] showed how the relationship between R_0 and r is determined by $g(\tau)$. Writing this relationship in the time-varying setting [134], we have:

$$\frac{1}{R(t)} := \int_0^\infty \exp\left(-r_t \cdot \tau\right) g(\tau)d\tau. \tag{2.3.3}$$

A key use of the (intrinsic) generation time distribution is for estimating time-varying transmissibility — via $R(t)$, often in renewal equation frameworks. A renewal equation describes how new infections at time t, $i(t)$, are generated from past infections, $i(t-\tau)$, at times $t-\tau$:

$$i(t) = R(t) \int_0^\infty i(t-\tau)g(\tau)d\tau. \tag{2.3.4}$$

The renewal equation framework is the basis for many popular approaches and software for estimating changes in pathogen transmissibility [137,139,148,149], and more generally can describe the births of newborn individuals in any structured population (e.g., modelling cell growth and fission) [150,151]. A key advantage is its inherent simplicity, relying on few modelling assumptions, while generalising to i) many modelling assumptions, from branching processes to compartmental models, ii) extensions to address challenges of epidemiological modelling (Sections 2.4 and 2.6), and iii) various stages of an epidemic [140,144,148,149,152].

However, a misspecified generation time distribution or incorrect modelling assumptions can produce biased estimates of $R(t)$ and, given the popularity of $R(t)$ for decision-making, potentially mislead decision-makers (e.g., if used to measure the effectiveness of interventions or current epidemic dynamics) [107,132,136,141,143,145,153]. r_t is more robust to misspecification of the generation time distribution, yet if the distribution is correctly specified, estimates of $R(t)$ can provide individual-level information and important implications for epidemic control, unlike r_t [134,142].

For mosquito-borne pathogens, there are additional complexities, induced by indirect transmission, the multiple stages required (Figure 2.1 B) for human-to-human transmission, and the many time-varying influences (environmental, biological, and human, Figure 2.1 C) affecting each stage of this transmission cycle [83,87,119]. For example, local mosquito adaptation, behaviour, and genetics can result in transmission heterogeneity across space and time [104,121,154,155]. Meanwhile, researchers have estimated temperature dependencies in the generation time distribution for dengue [80,156]. However, this derivation for the generation time distribution highlights a key outstanding challenge in estimating $R(t)$ for mosquito-borne pathogens; developing a probabilistic mathematical framework that allows more general space- and time-varying influences on the

transmission cycle, beyond purely temperature forcing. Such a framework should be able to handle uncertainties in epidemic quantities, extensions from laboratory to the field, and variations in climatic effects across space and time. However, even for directly transmitted pathogens, it is non-trivial to measure generation times [146], and mosquito-borne pathogens require further challenging data collection. We will explore these issues in Sections 2.4 –2.6.

2.3.3 Predicting epidemic trajectories

Given the concerns about model misspecification and practical data constraints, there are limitations to estimates of reproduction numbers and growth rates that are as accurate, timely, and relevant as is required for targeted public health policy that can respond to rapidly changing epidemic conditions [107,132,144]. These quantities are just a subset of the infectious disease modeller's toolkit for informing policy. Many further analyses can be complementary and valuable, including i) inference for latent variables or parameters that capture pathogen-specific dynamics (e.g., asymptomatic transmission proportions [4]) and drivers of transmission (e.g., climatic effects [79]), ii) nowcasting and other methods that correct for epidemiological data delays [113,114,157], and iii) predicting future epidemic trajectories (forecasting, scenario predictions and projections) [3,103].

We distinguish forecasting from projections by the horizon at which predictions are made; forecasts being short-to-medium-term (e.g., sub-seasonal and seasonal forecasts) and projections often being long-term (e.g., dengue projections for Europe by the end of the 21st century under different climate change scenarios [39]) [18]. The idea of probabilistic forecasting is to predict the future value of an epidemiological quantity, together with uncertainty. For example, the target quantity could be incident dengue cases and the uncertainty could be described using quantiles or samples of the predictive distribution.

Unlike scenario predictions, forecasts are unconditional predictions – these are a model's best probabilistic estimate of the future value of the target epidemiological quantity. A forecast generally does not predict future policy changes which could be influenced by the forecast, and instead integrates over the possible future changes in the world (i.e., scenarios) by using the model's assigned probability to each "what if" realisation/scenario [15,18,158].

Thinking of a more trivial example of a rainfall forecast informing a decision to take an umbrella, epidemiological forecasts should provide valuable information to decision-makers, where value is defined with respect to the decision-makers' ability to take timely actions (e.g., interventions to avert infections or deaths) while minimising associated costs (e.g., intervention costs) [3,18,159,160,161]. For example, forecasts of future dengue cases with longer lead times should be able to inform proactive vector control and public information campaigns ahead of time, rather than reactive insecticide spraying [159,162].

Although the accuracy and utility of an epidemic forecast is bounded above by i) the timeliness and representativeness of data [163,164,165] and ii) the unpredictability of the dynamical system [166], probabilistic forecasts are an important and increasingly common tool for informing longer-term seasonal planning for endemic diseases (e.g., dengue and seasonal influenza) and real-time outbreak responses for endemic and epidemic diseases (e.g., COVID-19 and Ebola) [3,159,167,168,169,170,171]. Forecasting competitions have found variable success of probabilistic forecasting models of incident cases [169,172,173], often suggested the potential for improvements in the reliability of forecasts, and concluded that infectious disease forecasting remains a poorly understood science [170]. For example, leveraging additional data streams and improving forecasting performance in key epidemic phases were recommendations from results of forecasting competitions for COVID-19, dengue, and seasonal influenza [167,169,170]. We will describe forecast modelling techniques, evaluation metrics, and the process of co-defining policy/research questions before and during infectious disease outbreaks in Sections 2.6 and 2.7 and Chapter 8.

2.4 Challenges for epidemiological analyses

Epidemiological modelling is often fundamentally about trying to elucidate the dynamical processes that produced the observed data. We would like our data to be timely, relevant, accessible, standardised and of high accuracy and resolution to allow modelling procedures to answer, with uncertainty, the questions of public health authorities[5,20,174,175]. However, the epidemiological surveillance data that we rely on are generally incomplete, noisy, and delayed, with quality issues caused by reporting, sampling, and testing[115].

These issues can therefore limit the utility of models for informing targeted and timely policies, and the difficulties are exacerbated in resource-limited countries affected by dengue[157,162]. Passive surveillance is the most common type of DENV surveillance, which is therefore based on generally symptomatic individuals presenting to health care facilities[176,177]. This creates issues for monitoring and modelling epidemic dynamics due to the high proportions of asymptomatic dengue infections. Furthermore, only an estimated 13.2% of symptomatic dengue infections were reported to surveillance systems across Southeast Asia[178], and similar underreporting is estimated for dengue incidence and hospitalisations in Brazil[7,179].

2.4.1 Reporting rates, representativeness, and standardised data

At a basic level, definitions of fundamental indicators can vary across space and time. For example, countries might report probable and/or clinically confirmed cases and can vary by their adoption of WHO definitions for diseases such as COVID-19 or dengue.[7,180,181,182]). Further issues arise for reporting rates due to i) underascertainment as infected individuals may not seek healthcare, ii) underreporting as healthcare authorities fail to report all cases, or iii) misdiagnoses (e.g., for dengue, if similar flaviviruses such as Zika or chikungunya co-circulate)[79,179,183,184,185]. These heterogeneities in reporting rates can occur within or between the geographies and time intervals of the reported data. All such issues highlight the need to work with local individuals and public health authorities who collect and/or co-ordinate data collection to better understand the data quality for the defined policy/research question and associated modelling task[105].

The causes of these heterogeneities are numerous, including inadequate access to healthcare, age/gender/race-related biases, time-varying self-reporting rates in a population, and/or the extent of epidemiological surveillance , all of which may be amplified during an epidemic[115,186,187]. These temporal and spatial heterogeneities can ideally be modelled explicitly to ensure that outputs are not confounded by temporal or spatial artefacts. For example, correction for underascertainment may be supported by representative seroprevalence data[117,122,188,189], while microscale epidemic dynamics may be unveiled by phylogenetic data[190]. However, these data are often unavailable in real time for many outbreak settings of the current thesis, especially during severe dengue outbreaks when surveillance systems can be overwhelmed. This motivates indirect modelling of heterogeneities, including expansion factors to estimate the true burden of dengue[7,178] or random effects in models of climatic effects[78].

2.4.2 Delayed and truncated data

In addition to underascertainment, underreporting, and misdiagnoses, we rarely if ever, know the exact time that an individual contracted an infection, even in individual line-list data[113,114,145]. Infection dates are censored, as infections, if reported, are often registered as new cases on the date of symptom onset or clinical testing. These issues are exacerbated for mosquito-borne pathogens, where mosquito infection prevalence and infection dates are generally unknown in real time due to insufficient, unrepresentative, and poorly standardised mosquito surveillance data[7,191].

Accurate representation of any epidemiological delay distribution is difficult due to censoring. The family of delay distributions includes serial intervals (times between the onset of symptoms in primary and secondary cases), generation times, and incubation periods (times from the onset of symptoms to infectiousness)[113]. Their real-time estimation is

further impeded by truncation (not all occurred events have been reported in the data yet), growing or shrinking epidemics, and/or insufficient real-time data availability. For directly transmitted infectious diseases, due to advances made since the COVID-19 pandemic, there are a growing number of methods to estimate delay distributions [113,114,145]. These methods are important because evolving epidemic dynamics can render both delayed data and associated model outputs out of date [175]. Noisy, delayed, and truncated data may appear to be theoretical concerns, yet the practical consequences of unknown totals can be substantial as the utility of model outputs for informing public health decisions is bounded above by their timeliness and representativeness [17,144,165,192].

2.4.3 Averages, heterogeneities, and communication

In epidemiology, knowing population averages helps us to understand individual outcomes, and vice versa [116]. This apparent paradox infiltrates our analyses, as many population average quantities have individual-level interpretations. For instance, reproduction numbers or generation time distributions are estimated as expectations across the population, yet are applied to a so-called "average" individual. Incubation periods are often measured at the individual level (e.g., for DENV [80]), vaccines are administered to individuals (e.g., for DENV [193]), and generation times can be measured using individual source-recipient pairs (e.g., for SARS-CoV-2 [146]), yet these are most often applied in a model to a population. The general observation is a manifestation of the dichotomy of "sick individuals and sick populations" introduced by Rose [194] in 1985 – aetiology demands that we understand both population-level and individual-level determinants.

The paradox is less counterintuitive when we consider that if the aim of modelling is to benefit the public health of individuals, then it is necessary to understand the overall dynamics of the epidemic to inform population and sub-population policy for individuals [3,3,4,5]. Indeed, heterogeneities are everywhere in epidemics – ranging from individual-level susceptibility and infectivity to spatially and temporally varying exposures [62]. Using the common example of modelling a pathogen-specific contact process, when employing population-level data, a modeller often assumes random mixing in non-network-based models. This often involves invoking a law of large numbers, yet heterogeneities are likely in human and mosquito populations due to human, biological, and environmental factors (Figure 2.1 C) [62,78,130,195,196]. Heterogeneities in susceptibility, exposure, or transmissibility can also be time-varying (e.g., time-varying contact patterns) and amplified by public health policy, the epidemic trajectory itself, and/or the environmental, human, and biological factors that have sustained or promoted epidemic growth (e.g., inequalities during the COVID-19 pandemic [21,22,197]).

A modeller will never be able to model all such heterogeneities. However, to reliably and equitably inform public health policy, it is important to consistently challenge and critique the following: the question- and setting-specific suitability, plausibility, and impact of modelling assumptions on results, as well as the level/scale and timeliness at which estimates are required for local policy [103,143,165,198,199,200,201]. Methods to address heterogeneities vary depending on the setting, the model, and the uses of model outputs. Methods often involve structuring a population by relevant characteristics/traits, invoking stochasticity in models, employing explicit spatial and temporal considerations, and/or using sub-population-specific data (e.g., socioeconomic factors in models or defining household reproduction numbers and generation time distributions) [62,78,116,130,137,146,200,202].

These heterogeneities and data quality issues (in reporting, sampling, and testing) are all contributors to the uncertainty that is associated with the outputs of any epidemiological modelling analysis [20,103,203,204]. The propagation and communication of the uncertainty through the modelling process, from inputs to final outputs and evidence synthesis, is a fundamental statistical challenge faced by modellers [103,205]. Communication has practical consequences for providing reliable information and maintaining the trust of local decision-makers and the wider public, and the dissemination of any modelling information requires clarity of audience – from scientific advisors to decision-makers to the wider public [105,160,203]. See Section 2.7 and Chapter 10 for further details.

2.4.4 Nonlinearities

Epidemiological data and modelling are plagued by another general phenomenon – the world is not linear. Nonlinear dynamical systems are complex and often poorly understood, yet surround us; in biology, chemistry, ecology, finance, neuroscience, and many other sciences[206,207]. The highly nonlinear processes of infectious disease outbreaks are generally modelled using limited and flawed knowledge of the epidemic[83,99,103,208]. These processes are perhaps most well known in the early exponential growth phase. During this period, the rate of change of new infections is proportional to the current number of infections. Understanding this exponential growth is crucial as it controls the timing of the epidemic peak and shape of the epidemic curve[15]. However, exponential growth is a finite process and cannot last forever as the susceptible population depletes. A modeller needs to monitor when this exponential growth phase ends to avoid an unrealistic model assumption and track the complex (e.g., chaotic or periodic) transmission dynamics for more advanced stages of an outbreak[17,62,209,210].

On the other hand, logarithms are ubiquitous and enable description of scientific phenomena in terms of magnitudes. Logarithms can also promote computational stability and accuracy when modelling dynamical systems. For example, logarithms enable greater precision for very small values such as probabilities. These theoretical and computational traits prompt the advised usage of logarithms in many analytical and numerical epidemiological settings[122,123,211,212,213].

More generally, nonlinearities can motivate functional approaches; mathematical, statistical, and/or deep learning techniques which can transform function spaces into spaces where dynamics may be closer to linear. Examples of techniques include kernel methods, normalising flows, and deep autoencoders. A nonlinearity common in mosquito-borne transmission is that of climatic exposures such as temperature which has nonlinear effects on mosquito biting rates, reproduction, development, and survival, and therefore on the observed human epidemic dynamics[80,82,83,214]. The nonlinearities include temperature optima and corresponding lower/upper bounds for either the virus or mosquito. These may be difficult to explicitly model while simultaneously propagating uncertainty i) from potentially unobserved variability in the data as diurnal mean temperature can impact DENV[215] and malaria[216] transmission, and ii) from the many other external modulating and interacting forces such as human water storage practices which can impact the availability of breeding sites[83]. Indeed, these issues are exacerbated by the estimated mosquito and viral traits being largely drawn from highly controlled laboratory experiments[80,82,88,217] and the lack of high-quality, large-scale mosquito surveillance data[83,191]. This increases uncertainty when extrapolating from laboratory-based results to observational studies in the wild with different mosquito population dynamics and genetics and different spatial and temporal scales[83,104,154,155,195,218].

All such nonlinearities suggest that modelling can benefit from i) additional data streams such as climatic, serological, and socioeconomic data for estimating environmental, biological, and human effects respectively[219,220], and ii) modelling approaches that can handle general temporal and spatial variations in mechanisms/relationships (mechanistic models[195,218,221]), make direct probabilistic inferences to quantify uncertainty (statistical models[78,79]), and/or can describe complex nonlinearities with many data-driven parameters (deep learning models,[222]).

2.4.5 Predictions, predictability, and governing laws

Unlike scenario predictions, a probabilistic forecast is the best estimate of the future value of an epidemiological quantity, together with uncertainty[15,18,158]. Considering simultaneously all future "what if" scenarios imposes uncertainty, which is compounded by incomplete knowledge of the present and past epidemic dynamics[103,131]. Unlike physics, chemistry, and other branches of science disciplines, there are generally no governing physical laws for dynamics in infectious disease outbreaks. For example, there is no analogue to Newtonian dynamics as epidemics are subject to heterogeneities, modulators, and confounders in transmission across individuals, space, and time. As modellers, the realised epidemic dynamics that we predict are both nonlinear and highly stochastic[99]; the coalescence of biological, human, and environmental influences

and their interactions (Figure 2.1 C).

Accounting for all such heterogeneities and disentangling the effects of different mechanisms, while making predictions of the future is difficult. This can limit the success of mechanistic-model-based forecasts, as observed for dengue and COVID-19 [169,223], especially when there are evolving model structures and/or local differences in epidemic dynamics. This problem can be viewed as a modeller having incomplete knowledge of time- and space-varying mechanisms and the corresponding data distribution shifts. In theory, this may induce limits to predictability in epidemic forecasting and suggests also focusing on alternative scenario predictions. Nevertheless, recent work indicates that the predictability limits are often beyond the time scale of single outbreaks for most infectious diseases [166]. These findings suggest that reliable short-term epidemic forecasting is likely difficult but possible, and this is supported by empirical evidence from recent studies for COVID-19 [170,172] and dengue [162].

2.4.6 Spatiotemporal obstacles

Epidemics unfold across space and time. Their dynamics have distinct spatial and temporal signatures. Spatiotemporal analyses are about modelling such dynamics across (often discrete) space and time.

Spatially, returning to heterogeneity, epidemic dynamics and drivers are i) most likely not spatially homogeneous *within* geographies especially in higher spatial scales/resolutions where nonlinearities are pervasive [104] and ii) potentially different yet interdependent between different geographies as human movement supports outbreak establishment, spread, and persistence [224,225,226,227]. Focusing on different spatial scales, Srivastava and Chowell [228] unveiled spatial heterogeneity in finer-scale COVID-19 growth rates in the United States (US) and found that averaging over these growth rates, by using coarser resolutions, blurred heterogeneity. Perkins et al. [218] captured the effects of spatial scales on mosquito-borne transmission as poor "mixing" (a typical modelling assumption) constrained the areas over which transmission occurred. Similar scale dependencies have been identified in other studies of dengue [195,229] and seasonal influenza [230]. More general spatial limitations encountered by epidemiological studies include boundary effects [231] and spatially varying confounding influences (e.g., mobility and climatic influences [232]).

It can also be unrealistic to assume that the risk of an exposure (e.g., climatic conditions) on subsequent incidence is constant over time (e.g., during a week/month), as this would not account for unmeasured variability in the aggregated data. On the other hand, higher-temporal-resolution surveillance data (e.g., daily reported cases) may have day-of-week reporting or testing effects. In general, temporal aggregation of epidemiological data involves integration over multiple possible epidemic trajectories [15]. Further temporal issues can present themselves when a modeller holds the temporal resolution fixed and models epidemic dynamics over multiple time periods (e.g., several months). For example, a climatic driver may have differential impacts on reported dengue case incidence over time due to time-varying human (e.g., water storage practices) or biological (e.g., serotype switch) influences. Ideally, researchers could estimate time-specific parameters, yet limited data and/or model structure constraints (e.g., non-identifiability) may prevent this [103,199]. Alternative indirect approaches include model fitting in trailing windows, model evaluation with sensitivity analyses, and cross-validation [148,233,234] (see Section 2.7). These approaches can help to determine the extent to which assumptions impact the reliability of outputs for the defined policy/research question.

To directly address such challenges, the modeller often needs high-resolution data and complementary data streams, alongside bespoke modelling techniques (e.g., metapopulation [235] or age-structured models [236]), to divide up time periods and/or geographies/populations. This can account for heterogeneities across space and time, yet increasing numbers of divisions and more data can also yield high-dimensional and computationally challenging data and model structures for simulation, inference, and prediction (Section 2.6). Of course, the spatial and temporal resolutions of available data can be inherent constraints of the model's abstraction of reality that may or may not limit the robustness of conclusions and/or local public health utility of spatiotemporal analyses [103,199,237,238].

Table 2.2: **Data streams for epidemiological modelling:** Examples, uses, and limitations of different data streams used in modelling epidemics. All data streams have their own individual strengths and limitations, and are therefore often used collectively in models to complement each other.

Data stream	Examples	Uses	Limitations
Epidemiological surveillance	Reported cases, hospitalisations, deaths	Central quantities for monitoring pathogen spread	Representativeness, delays, truncation, reporting rates
Demographic	Censuses, household surveys	Revealing population infection patterns	Outdated, spatial and temporal resolutions, representativeness
Climate	Temperature, precipitation, ocean indices	Quantifying climatic drivers of epidemics	Spatial and temporal resolutions, reanalysis data often reported without uncertainty
Human mobility	Flight, mobile phone, bus transportation	More realistic modelling of pathogen spread across space and time	Representativeness, limited availability, privacy issues
Genomic	Viral sequence, serotype information	Reconstructing transmission patterns	Representativeness, availability, financial costs
Seroprevalence	Cross-sectional or longitudinal serological surveys	Revealing landscape of population immunity	Resource-intensive, imperfect assays
Infection prevalence	Community prevalence surveys	Estimating infection prevalence during epidemics	Representativeness, self-testing, resource-intensive
Wastewater	Pathogen concentration or genomic sequences	Identifying transmission patterns and pathogens/variants	Spatial misalignment, dilution
Mosquito surveillance	Abundance, insecticide resistance	Understanding mosquito population and transmission potential	Limited availability, resource-intensive, sparse coverage (space and time)

Returning to the average-individual paradox of Rose [194], the ecological fallacy is a well-known phenomenon in spatial studies, suggesting the risks of making inferences at an aggregate level and applying them to individuals within the studied area without accounting for either individual-level variation or detail lost by averaging [239]. In particular, the Modifiable Areal Unit Problem (MAUP) is the potential changing of conclusions if the spatial aggregation level is varied, while the Misaligned Data Problem (MIDP) arises as the spatial data are analysed at different resolutions to the original/raw data [237,240]. These problems highlight why Diekmann and Heesterbeek [116] refer to heterogeneity as "the art of averaging". There will be no optimal spatial nor temporal scale for modelling epidemics, and public-health-focused analyses are driven by the available data and scale(s) relevant for the setting-specific policy questions of public health decision-makers [104,105,190,241].

2.5 Complementary data sources

As any infectious disease outbreak is the coalescence of biological, human, and environmental influences (Figure 2.1 C) [18,62], to reliably explain and/or predict local epidemic dynamics and inform decision makers in a timely manner, a modeller often needs complementary data sources beyond traditional epidemiological surveillance data [19,105,163,174,199]. Demographic data, often provided by censuses, household surveys, or geospatial population estimates (e.g., WorldPop [242]), paired with demographic science, are often vital for disentangling epidemic dynamics, drivers, and effects [243]. As shown for both dengue and COVID-19, understanding the underlying age profile of a population and epidemiological quantities can help to better understand important unknowns, including age-specific transmission patterns, clinical detectability, and disease-induced fatality rates. This understanding can then be used to inform design choices for model type, model outputs, and public health policy [243,244,245,246].

In regions that experience hyperendemic DENV transmission, there has been a historical concentration of infections in children and young adults, but the average age of infection is increasing. This has been attributed to shifting population structures, as well as changes to vector control and reporting and diagnosis [71]. Unveiling these signatures in DENV

transmission is important for more targeted interventions to reduce transmission, morbidity, and mortality, but it relies on age-stratified population and epidemiological surveillance data. These data offer the additional advantages of the age distribution of cases being more robust in terms of underreporting/underascertainment[244].

The increasing availability of high-quality, and often novel, data also aids modellers in achieving explanatory and predictive goals[15,19,247,248]. Obtaining more data (i.e., volume/scale) is no substitute for poor-quality data[249], and the more novel data streams of this section are often subject to many of the same uncertainties as epidemiological data[199]. However, these data can offer unique or complementary benefits. For example, equipping models with additional data, in type and volume, can help overcome or alleviate many challenges, including data quality issues (e.g., underascertainment[183,188,250]), statistical concerns about models (e.g., non-identifiability[188,251,252]), or basic barriers to inference or forecasting (e.g., insufficient data for robust or insightful analyses)[19].

Complementary data sources can also create new opportunities to explore completely new research avenues[15,105]. For example, in the early outbreak phase of the COVID-19 pandemic in the absence of up-to-date epidemiological surveillance data, crowd-sourced line-list case data and flight evacuation data from Wuhan were used to estimate early infection patterns in China[61,253].

The presence of mosquitoes is generally essential for the transmission of mosquito-borne pathogens. Mosquito densities can be monitored by many different types of mosquito surveillance[254], often from many different sources, ranging from citizens and individual researchers to public health authorities. These data can have different levels of representativeness and accessibility[255]. Real-time standardised measurements of the abundance of mosquitoes (e.g., numbers trapped per night) in different studied areas would be ideal but likely infeasible at such a large scale due to economic and logistical challenges[256]. Despite the challenges of standardisation and funding, there is an increasing number of surveillance data repositories[254]. Alternative modelling approaches have also been routinely used (e.g., VectorNet by the European Centre for Disease Prevention and Control, ECDC[257]) to generate vector suitability maps[258,259,260] that combine mosquito survey data with known covariates. More novel mosquito surveillance schemes include citizen science (e.g., mobile phone application reporting), WBE, remote-sensing-based surveillance, and metagenomic sampling of mosquitoes[261,262,263,264].

2.5.1 Human mobility data

For infectious diseases that affect humans, the introduction and transmission of the pathogen within and between populations generally requires the movement of people[224,225,226]. In an increasingly connected and urbanised world, short- and long-range human movements are often important for seeding events during the early outbreak stages, for the continued spread, and for the expansion of pathogens[61,77,238,265]. Human mobility often synchronises the epidemic trajectories across different geographies, providing regular flows of susceptible and infectious hosts (and vectors, where applicable) for sustained transmission[123]. Therefore, human mobility data, in their many guises, are useful for capturing spatiotemporal signatures of transmission and understanding how human movement affects epidemic spread[266]. Such data all have their inherent biases (e.g., sampling and censoring) yet can offer unique advantages for capturing different signatures of human movement[266,267], reflecting commuters, migrants, and international travellers. Using mobility data generally improves on model-based approximations of human mobility (e.g., gravity models) which can induce biases in analyses. The availability of these data streams in real time remains limited in many LMICs where dengue burden is high[238].

Ideally, these data allow modellers to describe the contact process in high spatial resolution and to make more accurate inferences or predictions about epidemics[163,245]. For example, aggregated human mobility data from mobile phones have been used to analyse many infectious diseases; from vector-borne diseases such as Zika and dengue to directly transmitted diseases such as COVID-19 and seasonal influenza[85,225,268,269]. During the COVID-19 pandemic, Apple and Google collected aggregated data from mapping applications on individual devices[270,271]. These open-access movement data facilitated spatiotemporal modelling of many unknowns – from assessing the effectiveness of public health interventions

to predicting future epidemic trajectories[61,269]. Similarly, in the UK, the COVID-19 mobile application of the National Health Service was downloaded more than 21 million times and opened new avenues for contact tracing[234,272]. Other studies have leveraged public transport data, commuter data, and flight data to quantify the influences of different types of human movement on epidemic dynamics[86,273].

For data collection, a vital component for future pandemic preparedness[19,175], the type of mobility data relevant to the decision-maker's needs can depend on the pathogen. For example, pathogens with shorter generation times demand short-term movement data[274]. From a causal perspective, incorporating an appropriate mobility data stream can be vital to ensure that model-based inferences are not confounded by movement between spatial units[232,274]. Equipped with such a diverse range of relevant mobility data (contingent on compliance with data governance requirements), there are opportunities for modellers to assess and predict pathogen spread through networks of individuals and/or populations. Working in this way, modellers can combine individual- and population-level data to provide public health authorities with a more realistic picture of how best to control pathogen spread[275,276].

2.5.2 Climate and weather data

Epidemics of infectious diseases exhibit temporal trends, predictable variability, and non-predictable variability (i.e., stochasticity)[123]. Disentangling these patterns from epidemiological surveillance time series data alone is often infeasible. Focusing on seasonality in mosquito-borne diseases, climatic and weather sensitivities in the vector and pathogen are well-established determinants of seasonal trends (Section 2.1.2 and[78,83,196,277]). We briefly draw the necessary distinction between *climate* and *weather*, as weather refers to short-term atmospheric conditions and climate describes the average weather over long time periods.

As the incidence and distribution of climate-sensitive[†] infectious diseases such as dengue continue to rise, the Big Data era has presented modellers with an increasing abundance of timely, accessible, and high-quality climate data. Climate reanalysis (e.g., WorldClim and ERA5 datasets[281,282]) combines historical weather station data with a physical model to provide high-spatial resolution estimates of ground-truth climatic conditions. Such datasets, while subject to their own often unknown uncertainties[283,284], provide opportunities to quantify the structure, timing, and intensity of relationships between climatic variables and subsequent incidence of new dengue infections/cases[78,79,214,246]. As climatic conditions are not the only driver of pathogen transmission, probabilistic statements about climatic influences are ideally made while simultaneously modelling, directly or indirectly (if unobserved), other influences and interactions (e.g., susceptible population recruitment or urbanisation).

Researchers are also aided by recent advances in short-, medium-, and long-term weather and climate forecasts, where advances have often been made by leveraging strengths of different disciplines (e.g., AI and physics)[285,286,287]. These weather and climate forecasts present new opportunities to improve the accuracy and local utility of climate-informed nowcasts and forecasts of dengue epidemics[53,159,162]. This is because integrating increasingly reliable weather forecasts can help tackle issues posed by delayed reporting, improve the timeliness and/or expand the horizon of forecasts, and ultimately allow public health authorities to take proactive actions[246]. Such actions may be decided on the basis of short-term (e.g., horizon of one month) and/or seasonal forecasts (e.g., the coming season)[246]. Meanwhile, infectious disease modellers can also borrow many of the methods that enabled these advances. These methods include interdisciplinary models, ensemble frameworks, and heterogeneous data assimilation[288,289].

[†]Throughout this thesis, we adopt the widely used description of an infectious disease as climate-sensitive (e.g.,[162,168,278,279,280]) if transmission is impacted by climate, weather, and environmental changes. To capture these sensitivities succinctly, we will often refer to climatic effects on epidemics that we estimate using climate reanalysis and epidemiological data.

2.5.3 More novel data streams

In recent years, many new methods for surveillance have been proposed. These provide new research opportunities and can correct for the biases of traditional epidemiological surveillance. WBE involves the collection of samples from wastewater sites that capture human excretions through urine or faecal matter. Due to the ability to detect viruses in the faeces of both symptomatic and asymptomatic infected individuals, WBE has enabled non-invasive surveillance of polioviruses and influenza A (H1N1) 2009 virus[290,291,292].

During the COVID-19 pandemic, as governments and public health authorities around the world sought to monitor the transmission and spread of SARS-CoV-2, WBE was proposed as a cost-effective surveillance method for early outbreak detection, pathogen importations, epidemic monitoring, and variant tracking[293,294,295,296]. WBE is far less developed for arboviruses compared to polioviruses or SARS-CoV-2[297]. As DENV is excreted in urine and faeces, several exploratory/pilot studies have investigated different uses of WBE, with variable success in the surveillance of DENV levels and vectors in wastewater[261,298,299].

WBE generally relies on another important surveillance method: genomic surveillance. At their core, genomic data provide fundamental measures of genetic diversity and have enabled researchers to enhance our understanding of how different pathogens function, evolve, and interact[300]. Genomic data, when paired with phylogenetic approaches, can help to unveil transmission patterns that are often missed by case surveillance data[301]. However, the COVID-19 pandemic underlined an urgent need to i) increase the publication of genomic sequencing results on open-access repositories and ii) support countries with limited sequencing capabilities[300]. There is a substantial global imbalance in available whole-genome sequences for DENV, with South America and Asia well-represented, unlike Africa (where dengue is endemic in some regions) and other continents[261].

To expand the scope of modelling analyses, researchers most often use genomic data alongside other data streams, such as epidemiological surveillance, mobility, climate, and wastewater data. For example, when used alongside wastewater or mobility data, genomic data helped modellers and authorities to track SARS-CoV-2 variants, their relative abilities to impact public health, and the underlying spatiotemporal epidemic dynamics[302,303,304]. Other studies have estimated $R(t)$ from genomic data to characterise the role of transmission heterogeneity and superspreading for different pathogens[305].

Community infection prevalence surveys are another data source that also use genomic data, and these surveys are carefully designed to allow population-level epidemic monitoring in real time. Such surveys also enable investigation of past epidemic dynamics and drivers such as vaccination, variants, and interventions[306,307,308,309]. While relatively rare during outbreaks due to logistical and economic barriers, these surveys combine individuals' self-testing results with epidemiological/genomic data, demographic data, and statistical methodology to estimate the percentage of a population currently infected with a virus. Community prevalence surveys were used in England during the COVID-19 pandemic to obtain nationally and sub-nationally representative estimates of SARS-CoV-2 prevalence (see Chapter 3)[306,307,308,309].

2.5.4 Seroprevalence data

Understanding the immunity and susceptibility landscape is critical for epidemic control – to understand who is at risk, why pathogen spread occurs, and where interventions should be targeted[310].

A serological survey, also known as a serosurvey, can help provide such understanding and address imperfections of traditional epidemiological surveillance data[310]. In a serosurvey, individuals are tested for antibodies of a pathogen, and individual-level seropositivity (determined if the level of antibodies exceeds a threshold) can then be used to investigate many important unknowns in epidemics including incidence and prevalence trends and population-level immunity/susceptibility profiles[189,310].

Age- and/or time-stratified serosurveys allow modellers to derive a measure of transmission intensity – the Force

of Infection (FoI), which is the rate at which a susceptible individual becomes infected per unit of time[244]. This reconstructed FoI can be used to estimate the historical reliability and representativeness of case surveillance data[244,311,312] and of any subsequent model-based inferences. This is especially useful for dengue due to policy challenges posed by high asymptomatic and underreported transmission rates in endemic countries with limited resources. Incorporating serosurvey data can also help to enhance reliability of model-based inferences or forecasts by addressing nonlinearities in relationships (e.g., climate-infections) induced by rising/declining levels of population immunity[220]. However, similar to the interdisciplinary approaches of this thesis (multiple models, perspectives, and data streams), Huang et al.[313] highlight that serosurveys have their own imperfections (e.g., assay sensitivity) and it is important, for the accuracy of FoI and immunity profile estimates, to compare the inferences derived from different types of serosurveys (longitudinal and cross-sectional) and age-stratified case data.

Although they provide important immunity information for the shapes of epidemic curves, serosurveys are regarded as less useful for predicting future epidemic dynamics[314]. Serosurveys of DENV have also historically been limited in temporal and spatial coverage due to their resource-intensive nature, yet these surveys are increasing in number[315]. Similarly, a recent multipathogen dashboard (ArboTracker[316]) for arboviruses aims to address the twin issues of outdated and under-representation in serosurveys. ArboTracker builds on advances from the COVID-19 pandemic (SeroTracker[317]) and acts as a updating review which provides easily accessible seroprevalence data. Meanwhile, recent software developments, such as serofoi[318] and RSero[319], provide user-friendly tools to model pathogen dynamics from serosurvey data more easily.

While data availability in terms of their coverage (e.g., for time-varying FoI estimates) and inclusion of external factors (e.g., concurrent or preceding climatic conditions) remain key limitations of many dengue serosurveys[315], their multiple uses[189,314], increasing availability[315], and enhanced usability[316,318,319] indicate the expanding future avenues for research of dengue and other infectious diseases.

2.5.5 Raw data versus derived quantities

We close this section by drawing the distinction between raw data (direct measurements collected by hard-working professionals) and quantities/estimates derived from raw data. Although data themselves are subject to many challenges (Section 2.4 and[105]), derived quantities are further dependent on the data processing adjustment or modelling assumptions, and uncertainty in such estimates should ideally be preserved where possible. We will raise this distinction where relevant throughout the thesis and discuss their implications in Chapter 10.

As examples of derived quantities, the distinction applies to i) wastewater concentrations for SARS-CoV-2 which were reported in the UK post normalisation of wastewater flows[302] (Chapter 3), ii) survey-based prevalence estimates for SARS-CoV-2 which can be subject to weighting schemes for different demographics and regions[309] (Chapter 3), and iii) climate reanalysis estimates which have been spatially and temporally interpolated (e.g.,[281], Chapter 4).

2.6 Modelling infectious disease outbreaks

When modelling the dynamics of an outbreak[†], we can use data in many different ways. Guided by the policy/research question (Figure 2.3 A) and available data, different approaches chosen by modellers will encode different epidemiological, mathematical, and/or statistical assumptions (Figure 2.3 B)[62,103,131,160,241].

Any such model, in the absence of complete information, will always be an abstraction that must make appropriate simplifications about the observation process and the true, latent processes that produced the observed data. This creates inherent modelling trade-offs between three potentially conflicting elements: accuracy, transparency, and flexibility[122].

[†] In this thesis, we only model urban and rural transmission cycles and do not consider enzoonotic/forest transmission cycles[320].

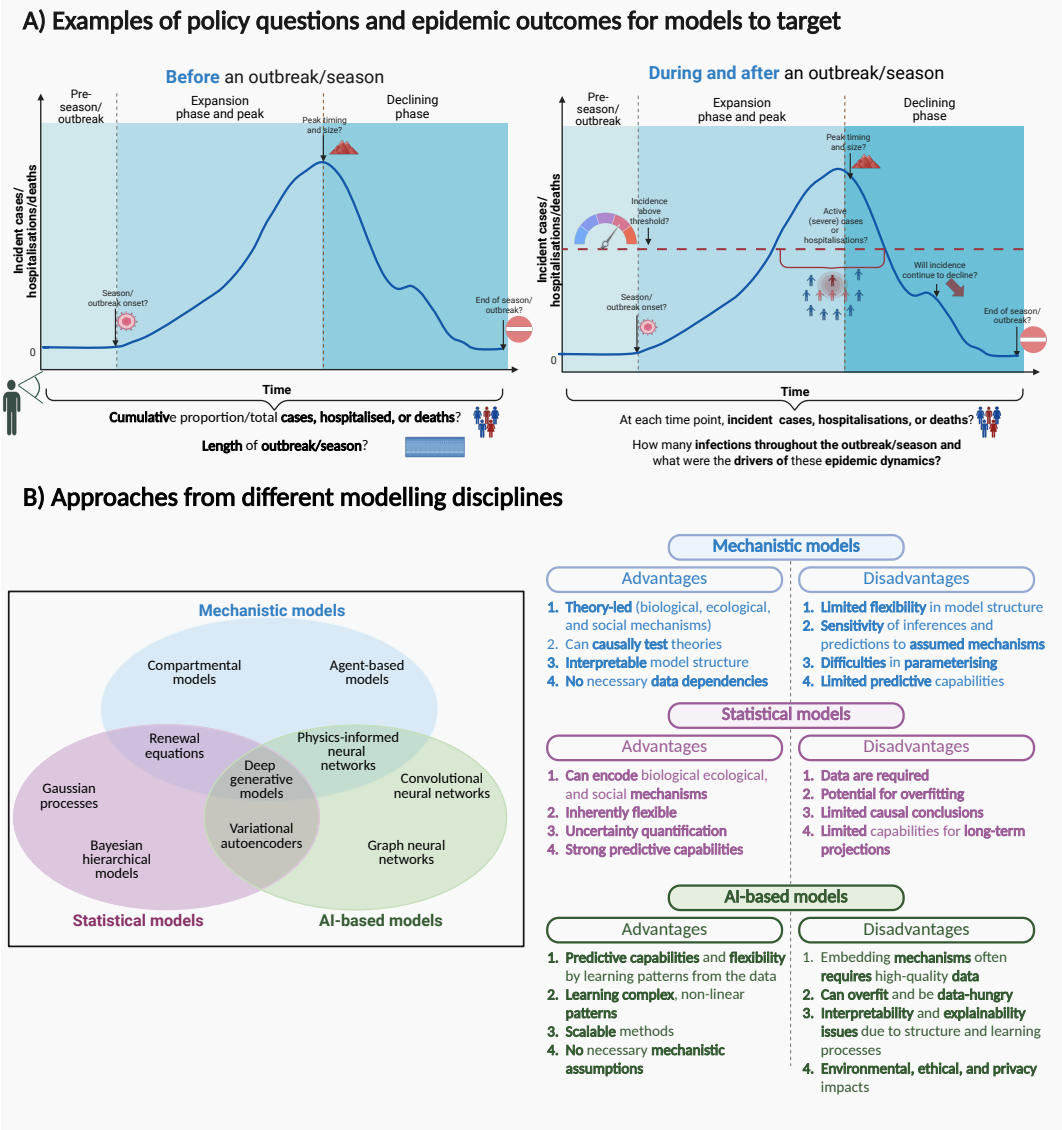


Figure 2.3: **Modelling approaches to answer different policy/research questions**. **A):** The modelling of this thesis is motivated by answering different policy/research questions, which vary across space and time. The epidemic curves (discussed through the lens of forecasting in Chapter 8) are labelled here with a subset of the many questions that epidemiological modelling targets before, during, and after seasons/outbreaks. **B):** A simplistic taxonomy of modelling classes/disciplines, where classes are not distinct and models can belong to one or more of the classes depending on their structure and assumptions. Four key advantages and disadvantages of each modelling class are also presented to highlight where such models are more or less suitable. This visualisation is complemented by the details of modelling procedure for public health policy in Figure 2.2 and of model evaluation and evidence synthesis in Figure 2.4.

Adopting a discipline-based and slightly reductive approach to the taxonomy of epidemic models (Figure 2.3 B), we separate the modelling classes/disciplines into i) mechanistic, ii) statistical, and iii) AI-based (machine learning and deep learning) models. Each class generally differs in the extent to which their models aim to mimic dynamical processes of transmission and in how they handle uncertainty. These broad classes are not mutually exclusive, as a modelling approach can belong to one, two, or all of the classes. For example, semi-mechanistic renewal equation models combine mechanistic and statistical modelling strengths, while the inference and/or simulation/optimisation of mechanistic and AI models can often rely on statistical techniques.

The following descriptions are complemented by discussion throughout the thesis, are non-exhaustive, and focus

on modelling epidemic dynamics. Other approaches such as phylogeographic methods can combine phylogenetics with population genetics to uncover spatiotemporal spread of pathogens [301,321], and risk mapping methods can combine incidence, occurrence, and seroprevalence data to jointly infer pathogen risk [322].

2.6.1 Mechanistic models

History of mechanistic modelling

Mathematical models for infectious disease outbreaks date back to the eighteenth century from Daniel Bernoulli's model of smallpox in 1760 [323]). However, it was Ronald Ross who, building upon the pioneering work of scientists (e.g., the germ theory of infection of Louis Pasteur and Robert Koch [324]), developed the first known mechanistic models of transmission for malaria between 1908 and 1921 [325]. Before Ross' seminal modelling work for mathematical epidemiology, a doctor called Patrick Manson demonstrated that mosquitoes transmitted a blood-borne pathogen [326]. Charles Laveran, a French army officer, had also found malaria parasites in the blood of human patients. These findings lent themselves to theories of mosquito-borne transmission of malaria [121,325]. Consistent with common modelling objectives today, Ross' model development demonstrated a strong public health focus and was influenced by candidate mosquito control measures to eliminate the ancient disease of malaria [324]. Ross' final models, co-developed with Hilda Hudson across 1916–1917 [327,328,329], were a precursor to the now-ubiquitous susceptible-infected-removed (SIR) model of transmission (see below). The SIR model was a simplification of the more general Kermack and McKendrick model developed by William Ogilvy Kermack and Anderson Gray McKendrick [121,330]. In the past century, there have been countless advances which build on these initial mechanistic models to improve our understanding of complex epidemic processes for many infectious diseases [117,122,330].

It is common to focus on the contributions of individual modellers/researchers themselves (e.g., Ronald Ross), yet we emphasise in this thesis that their scientific research and historical advancements often relied on collaborative efforts (e.g., Hilda Hudson) and field studies of infectious disease dynamics alongside local participants (such as malaria patient Abdul Kadir in Ross' seminal studies) in LMICs (e.g., McKendrick, Ross, and Macdonald all studied malaria in Sierra Leone) [121].

These mechanistic/biomathematical models of transmission, progression, and control differ from historical precursors (e.g., descriptive methods of John Snow and William Farr for cholera [331] and smallpox [332], respectively) by making explicit *a priori* assumptions about the epidemiological mechanisms that drive infection dynamics (i.e., *how* transmission occurs). In doing so, a mechanistic model uses a quantitative perspective. For Ross, it was describing the dynamical system of malaria transmission in populations. This quantitative perspective can then be used to explore and/or infer key epidemiological parameters/patterns [333]. This creates in silica environments where a modeller can test counterfactual situations (i.e., "what if" questions) such as the effects of interventions without performing real-world experiments [334]. For mosquito-borne pathogens, the Ross-Macdonald theory has long been the pre-eminent basis for mechanistic models. There are now many variants of these Ross-Macdonald models, yet at their core, they revolve around key concepts of mosquito spatial movement, mosquito survival, and biting rates. These concepts are summarised in model parameters of R_0, vector carrying capacity per human, and mosquito density required for transmission [104,121].

Mechanistic modelling at different scales

Mechanistic models can vary in their assumed mechanisms[†] and scales [122], both of which impact their computational and theoretical complexity. Different mechanisms can capture different times taken between infection events (e.g., the generation time) or different host/vector transition rates (e.g., birth and death rates). Model scales can be within-host, individual, or population levels. Within-host models describe dynamics at a microscopic level and are commonly used for

[†] In this thesis, we use mechanism to refer to epidemiological mechanisms, as opposed to other mechanisms (e.g., statistical mechanism).

quantifying regulators of viral load dynamics[335]. While not explored in this thesis, these models are crucial for dengue in exploring the impacts of the currently limited dengue vaccines[336].

Agent-based models (ABMs) are bottom-up, microscale models that mechanistically simulate epidemic dynamics at a granular level. Agents, usually individuals, are assigned specific attributes which allows a modeller to test hypotheses about different mechanisms and interactions in complex epidemic processes[122]. ABMs, often termed individual-based models, allow for stochastic events such as transmission and capture the individual-level interactions with an epidemic. Returning to Rose's paradox of studying individual-level dynamics and outcomes to inform population-level policy[194], ABMs can help to understand the complex interplay between individual-level and population-level epidemic dynamics[337]. This can help resolve the challenges posed by feedback loops between individual behaviour and population infection dynamics. For example, understanding contact processes between individuals helps to understand the potential for the spread of epidemics across space and time[62,108,338]. Different ABMs have informed epidemic control strategies for smallpox[339], tuberculosis[340], and H5N1 influenza[341]. During the COVID-19 pandemic, modellers used ABMs to estimate the potential effects of vaccination, testing strategies, and containment measures[342,343,344]. Computational complexity (e.g., often requiring model-specific optimisations) and barriers to parameter inference (e.g., insufficient individual-level covariate data)[108] have historically limited widespread use of ABMs for real-time epidemiological modelling tasks[122,337]. However, deep neural network architectures, approximate Bayesian computation (ABC), and variational Bayesian inference have all provided new opportunities to approximate ABMs[345,346,347].

Mechanistic modelling is most often performed at a simpler population level. Compartmental models subdivide the population into distinct categories/bins known as compartments. These models deviate from ABMs by treating the compartments as homogeneous groups based on one or more characteristic such as infection status and age. Compartmental model structures were first introduced in the seminal work of Kermack and McKendrick[330], where the general form was a PDE framework for the SIR model. This framework structured the infected population/compartment (I) in terms of infection age (time since infection) and time, while the susceptible (S) and removed populations (R) were modelled more simply over time without any age structures. Further extensions include separate exposed (but not yet infectious) and infectious compartments, and demographic factors such as birth, migration, and death rates. The benefits of compartmental models (Figure 2.3 B) include their i) inherent simplicity (often as a system of ordinary differential equations – ODEs) and interpretability, ii) flexibility to a wide range of infectious diseases and decision-making questions, and iii) ability to explicitly consider different causal mechanisms of the transmission cycle and therefore make important causal inferences to inform public health policy[108].

The fitting of increasingly complex mechanistic models has been aided by the advent of big data and new data streams[20]. In some settings, the imposed structural assumptions of mechanistic models, such as *a priori* assumed mechanisms, fixed parameters, and less flexible structures, can offset some of their advantages (Figure 2.3 B)[348].

Deterministic and stochastic mechanistic modelling

Deterministic compartmental models form the basis for analogous stochastic models such as stochastic SIR models. Both deterministic and stochastic compartmental models belong to a wider class of state-transition models. While deterministic models are generally written as a system of ODEs, stochastic models are often written as continuous-time Markov chains (CTMCs) or stochastic differential equations (SDEs). The CTMC setup uses stochastic counting processes for each compartment and forward and backward Kolmogorov differential equations to derive infinitesimal transition probabilities[349]. SDEs similarly use Kolmogorov differential equations and modify the deterministic systems to introduce stochastic fluctuations in compartments.

Stochastic models can be important due to a modeller's imperfect knowledge of epidemic dynamics[103] and the fact that the establishment, expansion, and persistence of almost all infectious diseases are subject to some stochasticity (e.g., early

stages of the COVID-19 pandemic[350]). Using deterministic compartmental models can often be justified when the number of infected individuals grows by a law of large numbers argument[62]. However, stochastic epidemic models are increasingly important when there are small population counts (e.g., small populations within metapopulation units), low infection counts, or variability that affects epidemic outcomes and spatial spread (e.g., transmission, recovery, or environment)[122].

With increasing climatic variability[351,352,353], urbanisation[78,79], and globalisation[77,354], DENV is moving into areas that have never experienced DENV infection before[355]. These populations, often on the fringe of historically at risk areas, are likely to experience lower infection counts and more sporadic outbreaks as climatic conditions become increasingly suitable for transmission[351,356]. Such small numbers indicate the need for stochastic systems, representative of a general theme that epidemic dynamics are always subject to both demographic (i.e., individual-level) and environmental stochasticity[122].

There are three main effects of stochastic dynamics in a mechanistic model: i) the system can be pushed away from a deterministic attractor, ii) chance extinctions of the epidemic are possible, and iii) variances and covariances can change the deterministic behaviour[122]. Deriving mean behaviour and higher order moments of a stochastic system is non-trivial, as the behaviour of the deterministic model can (e.g., for dengue[357]), but does not necessarily, approximate/mirror the mean behaviour of the stochastic system. Methods for deriving such behaviour include analytical approaches (e.g., moment closure) and numerical simulations (see Chapter 7).

Spatial structures in mechanistic models

While embedded in many mechanistic and semi-mechanistic models (e.g., renewal equations), assumptions of perfect mixing for hosts and/or vectors within and across spatial units can be unrealistic (Section 2.4). Mechanistic models can be extended to deal with potentially high-dimensional spatial structures. The type of spatial model generally depends on the host (and vector, if applicable), available data, and pathogen in question[200].

For example, a metapopulation model can consider connections between spatial units via a graph structure, where spatial units share common features[122,358], and can further incorporate a hierarchy of increasingly large spatial units (e.g.,[221]). The weights between spatial units can be time-varying and represent simple binary neighbours or connectedness between regions determined by mobility data and/or approximations[122,235,235,359].

Network models are related, more computationally intensive models that consider individual-level transmission alongside contact network data or graph theory. In network models, nodes are individuals and edges are contacts. These models can accommodate a wide array of network structures including different degree distributions, mixing properties, and clustering traits[360]. Often, network models reflect the individual-average paradox[194], as capturing the individual-level infection behaviour allows for describing epidemic dynamics at the population scales (e.g., for the interacting impacts of dengue vaccination and vector control[361]) relevant for public health policy[338].

2.6.2 Statistical models

Statistical modelling also represents a broad class of frameworks which can differ more (e.g., nonparametric models) or less (e.g., semi-mechanistic models) substantially from mechanistic approaches (Figure 2.3 B). The common denominator between all statistical models is that the model is fitted to data using an inferential framework.

Renewal equations

Bridging the gap between fully statistical and mechanistic models, semi-mechanistic models involve statistical estimation of parameters within constrained epidemiological mechanisms. Although semi-mechanistic models predate the COVID-19 pandemic (e.g., for Ebola Funk et al.[362]), their usage has expanded in recent years, leading to new methods for modelling the dynamics of endemic[363] and epidemic[148,364,365,366]) diseases. Renewal equations (eq. (2.3.4)) are often used to

incorporate epidemiologically plausible mechanisms. Used in many branches of science, from actuarial science[367] to demography[150], renewal equations entered the world of mathematical biology and infectious disease modelling through McKendrick von-Foerster equations[368]. Such equations have been embedded in Bayesian smoothers and generative Bayesian hierarchical models that estimate $R(t)$, the effects of non-pharmaceutical interventions, and other key indicators of epidemic dynamics[59,148,149,364]. Unlike statistical time series methods, as a semi-mechanistic method, renewal equations allow the modeller to infer the potentially causal effects of interventions and local population behaviours on transmission dynamics (e.g., via $R(t)$[59]), and compared to mechanistic models, help avoid making potentially strong a priori assumptions.

At their simplest, renewal equations assume no population or spatial structure and therefore assume all individuals in populations mix homogeneously and no infection importations occur. However, renewal equations can be extended to include heterogeneity in transmissibility[369], contact patterns[370], and infection importations[129,140,371,372]. Practically, renewal equations are often implemented in discrete time to align with data and their temporal resolution, for which the analogue to the continuous eq. (2.3.4) is:

$$I(t) = R(t) \sum_{\tau=1}^{t-1} w(\tau) I(t-\tau) \tag{2.6.1}$$

where $I(t)$ is the incidence of new infections at time t, calculated as a sum of past infections $I(t-\tau)$ weighted by the probability distribution for the (discretised) generation time $w(\tau)$ and the instantaneous reproduction number $R(t)$. Applications of renewal equations sometimes use incident reported cases only (and therefore do not account for reporting rates or reporting delays) and rely on serial interval estimates to approximate the generation time distribution[139,140]. Otherwise, case reporting rates can be inferred jointly alongside $R(t)$ in a renewal equation framework (e.g., by using deaths data[148]) and generation time estimates can be derived from infector-infectee pairs (e.g.,[146]). Further extensions and alternatives can address delays and truncation in epidemiological data. These methods can, but do not necessarily, involve complementary data streams (e.g., wastewater[373] or hospitalisation[366] data). For example, recent methods include estimating $R(t)$ from genomic data, instead of case data, alongside a branching process model (a more general renewal process)[305] and using Bayesian joint inference methods alongside retrospective updates to case time series[144].

Modellers generally incorporate stochasticity in renewal equations by assuming a likelihood for the observed case/infection data. The likelihood is a function proportional to the probability of the data. The likelihood in renewal equations is often a Poisson distribution to accommodate integer-valued data or a Negative Binomial distribution to further accommodate overdispersion[15,130,139]. These remain largely arbitrary and convenient modeller choices and may be restrictive. A Poisson likelihood does not account for overdispersion, and a Negative Binomial likelihood assumes a quadratic relationship between the mean and variance. This Poisson/Negative Binomial noise remains an unsolved limitation of popular frameworks based on renewal equations[141]. A further limitation is that the renewal equation theory has been designed for directly transmitted infectious diseases[137,139,374]. The corresponding theory has not been developed for vector-borne diseases that experience multiple stages of a human-to-human transmission cycle, each of which are often subject to time-varying and interacting conditions (Figure 2.1 B – C). The few statistical frameworks for estimating time-varying reproduction numbers for dengue have used alternative case (as opposed to instantaneous) reproduction numbers[375] or relied on time-invariant parameters for the generation time distribution[376,377].

Bayesian nonparametrics

Bayesian nonparametrics are another flexible class of models that can reduce the set of *a priori* assumptions. It combines Bayesian inference with nonparametrics and can handle complex data, while removing potentially inappropriate distributional assumptions by using a potentially infinite number of parameters[378]. Recent uses include estimating $R(t)$ in a piecewise-constant manner (in a Poisson renewal model) without specifying *a priori* the change points[379].

Gaussian processes (GPs) are a rich and inherently flexible family of Bayesian nonparametric approaches for regression and classification tasks. A GP is a probability distribution over functions and mathematically defines a collection of random variables where any finite number of them has a Gaussian distribution[380]. As GPs suffer from computational barriers ($O(n^3)$ computational complexity)[380], approximate GPs (e.g., Hilbert space methods[381]) are often required. These approximate methods have been used for many infectious disease modelling tasks, including inference of key parameters[148,188,382]), clustering/spatial risk maps[383]), small-area spatial modelling[384]), and nowcasting and forecasting[385,386]).

Bayesian hierarchical models

Bayesian hierarchical (or multi-level) models represent a very general framework, including both linear and nonlinear models[387]. These models are used widely across many branches of science, including ecology[388,389], gene expression[390], and infectious disease modelling. The framework assumes a nested/hierarchical structure, where the model is built at each level of the hierarchy. Using relevant data, the model parameters are estimated for all levels simultaneously, which accounts for group-level variations while estimating the overall fixed effects[391]. For epidemiological modelling, these hierarchical models may or may not include an assumed epidemiological (human, biological, or environmental) mechanism. More generally, Bayesian computation most often requires two key steps: computation of i) the posterior distribution(s) $p(\theta \mid y)$ and ii) the posterior predictive distribution(s) $p(\tilde{y} \mid y)$[391], where θ is the set of unknown parameters, y are the observed data (e.g., newly reported cases), and $\tilde{y}$ are new data.

There are many benefits of a Bayesian hierarchical framework, including i) its flexibility (extending to a wide range of models), ii) the wide applicability of the nested structure (e.g., to areal units or individuals), and iii) the partial pooling (estimates can borrow strength and allow uncertainty across space and time, and improve posterior inferences)[237,391]. As a Bayesian approach, a unique advantage is that the posterior distributions allow direct probabilistic statements about epidemic dynamics and epidemiological quantities. This is advantageous due to the lack of governing laws in outbreaks and the highly imperfect available data which are insufficient to capture all relevant mechanisms with certainty. For example, for dengue in Thailand, Huang et al.[236] demonstrated the advantages of pairing a Bayesian hierarchical framework with population and dengue case data to disentangle, together with uncertainty, the roles of several influences (e.g., changes in the age structure of the susceptible population and age-specific reporting) that increased the age of reported dengue cases.

When working with time series data during an epidemic, effects may i) take time to materialise, and/or ii) be nonlinear (see Section 2.4). For dengue, due to climatic sensitivities and human modulators, these delayed, nonlinear patterns can arise in relationships between climatic variables and incidence of new human disease cases. The nonlinear relationship, may be caused by accelerated mosquito larvae development, faster reproduction rates, greater mosquito survival rates, reduced extrinsic incubation period of DENV, heightened mosquito biting activity, and/or faster viral replication within the mosquito[80,81,83,217,392]. The delay in the climatic effect can be due to the mosquito life cycle (Figure 2.1 A), the incubation period of DENV, and/or the time taken for climatic conditions to influence larvae development[80,393]. These delayed, nonlinear relationships could further vary across space and time due to different human or biological factors.

Modelling these complex relationships is central to analyses for the effects of various influences on dengue transmission. These analyses often rely on Bayesian hierarchical models of dengue incidence alongside distributed lag nonlinear models (DLNMs). DLNMs are based on a cross-basis: a bi-dimensional functional space consisting of a combination of two sets of basis functions that represent the relationships with the response in the dimensions of the predictor and lags, respectively[394]. A DLNM simultaneously specifies i) nonlinearity in the exposure-response dimension and ii) delayed effects in the lag-response dimension. DLNMs in Bayesian hierarchical models have allowed modellers to probabilistically quantify the timing, structure, and intensity of human, environmental, and climatic influences on reported dengue cases in different countries (including Barbados[214], Brazil[78], and Vietnam[79]).

Statistical time series methods

Approaching epidemiological modelling from a purely statistical time series perspective can be a natural approach, particularly for short-term predictive tasks. This is due to the i) temporal structure of the data as surveillance and weather/climatological data, at a basic level, are often simply temporal measurements, ii) temporal patterns of variables including seasonality, autoregressive behaviour, or heteroscedasticity, and iii) opportunities to flexibly forecast future epidemic trajectories with relatively few assumptions (compared to e.g., a mechanistic model)[395]. An example of a statistical time series model is a Seasonal Autoregressive Integrated Moving Average, SARIMA, model

Time series methods do not track individual or population transitions between states, make fewer assumptions, and explicitly preserve temporal dependencies in the data. This means that a modeller is not exposed to the same risks posed by potentially invalid model structures or epidemiological assumptions (e.g., after data distribution shifts) in mechanistic models. Time series models such as SARIMA models are often used for benchmarking performance of more complex forecasting models. In the most well-known forecasting competition for predicting dengue cases, mechanistic models were generally outperformed by a simple SARIMA model[169].

Despite these predictive strengths, unlike mechanistic/semi-mechanistic models, an important limitation is that statistical time series models do not provide in silica environments for investigating important policy/research questions such as the causal effects of interventions on epidemic dynamics[59,334].

2.6.3 AI: Machine learning and deep learning models

Overview of model structures and training

For both endemic and emerging infectious pathogens, the increasing abundance of data type and volume (see Section 2.3) has advanced the development of data-driven machine learning and deep learning approaches to epidemiological modelling[53,396,397]. Throughout the thesis, we will consistently use a subtle distinction between machine learning and deep learning; defining deep learning as the subset of machine learning that uses artificial neural networks[398]. We use this distinction because machine learning itself can be a very broad class and is often defined as the computer-based process that improves its performance for a specified task by experience of the task[348,399]. Both machine learning and deep learning fall under the umbrella of AI and, in the context of this thesis, generally frame epidemiological modelling as a predictive task. Inference can be performed in such models through prediction. AI-based modelling involves learning of underlying patterns using loss minimisation algorithms to optimise model weights or analogous hyperparameters. The loss minimisation can be done with standard optimisation techniques (e.g., L1 or L2 loss) and/or newer regularisation techniques (e.g., dropout or batch normalisation) that help guard against overfitting[398]. Physics-informed machine learning (PIML) models often involve custom-designed loss functions which allow physics-based regularisation[287].

As the training of such models historically relied on sourcing large volumes of high-quality (representative and standardised) data which are often unavailable for epidemics, their applications to infectious disease modelling were limited for a long time. More recently, many of the AI-based methods have been developed specifically for directly transmitted diseases and most of them for COVID-19 due to the wealth of available data. This means that there are research gaps in the generalisability of such methods for vector-borne diseases and for resource-limited LMICs[289].

AI models in epidemiological modelling

AI-based models can differ from mechanistic models by not necessarily requiring any epidemiological mechanisms. Meanwhile, AI-based models can differ from statistical models by learning underlying patterns by optimisation without forming prior data assumptions, deriving functional expressions for parameters, or making inferences about relationships between variables.

However, AI-based models are not necessarily distinct from statistical or mechanistic models (Figure 2.3 B) as they can, for example i) embed epidemiological mechanisms (e.g., physics-informed neural networks – PINNs using renewal equations[400]) or statistical structures (e.g., foundational time series models[401]), ii) augment mechanistic and statistical models (e.g., surrogate modelling), and iii) be generative models that target the joint distribution (e.g., deep variational autoencoders, VAEs)[289].

External to epidemiological models themselves, machine learning and deep learning techniques can also assist in the implementation of mechanistic and statistical models by i) providing diverse, high-quality data such as weather forecast or genomic surveillance data[286,321,402], ii) assisting preprocessing and exploratory analyses (e.g., clustering or dimensional reduction[403,404]), and iii) constraining function spaces (e.g., plausible mechanisms for mechanistic models[53]).

In line with the ambition of the current thesis, researchers have often advanced the field of AI by integrating the advances of other scientific disciplines, including mathematics, statistics, neuroscience, information theory, physics, control theory, operations research, and many other fields[398,405,406]. These advances have been made across the design, training, implementation, and interpretation of models.

AI model types

Historically, machine learning for epidemiological modelling often involved techniques such as random forests, gradient boosting, and support vector machines. For both COVID-19 and dengue, these traditional models have been used for regression and classification tasks[407,408].

In recent years, more complex deep learning architectures, often called black-box algorithms due to their structural and learning complexity (e.g., many parameters), have made it possible to i) handle larger volumes and more variety in data, ii) learn complex, nonlinear underlying patterns, and iii) perform automatic feature extraction. These models include graph neural networks (GNNs), long short-term memory (LSTM) models, transformer models, and convolutional neural networks, which have been successful for monitoring and forecasting epidemic dynamics of both COVID-19 and dengue[86,222,397,409,410,411].

Attention mechanisms are often used in deep learning models of epidemics. Attention mechanisms aim to focus model training on the most relevant parts of input data. Alongside the use of potentially time-varying graph structures, this allows modellers to capture spatiotemporal signatures of epidemics[410,412].

Addressing the data-hungry nature of deep learning models, transfer learning has helped the development of foundational time series models, as these models have been pretrained on potentially billions of data points. Although underexplored for epidemiological modelling, such models could provide new avenues in data-constrained epidemic environments[53,289].

AI, generalisability, and explainability

Despite the rapid development of such methods, the advancement of computational resources, and the increased availability of sufficient data, complex model architectures and learning processes can mean that the benefits of AI-based models can come at the expense of a loss of explainability and interpretability[413]. This is the difficulty in explaining the model's inner workings and rationale behind its decision-making, and arises due to a model's black-box predictive structure. Furthermore, these models can, but do not necessarily, suffer from overfitting and lack of generalisability (Figure 2.3 B), which contrasts with generalisable and adaptable mechanistic PDE-based systems[348,414]. These issues are in addition to the increasing environmental and social impacts of AI models (see Chapter 10 and[415,416]).

In the last decade, many explainable AI (XAI) methods have been proposed to improve interpretability, understandability, justifiability, and transparency[417]. Explainability metrics are useful ways to better understand complex models, yet they cannot guarantee the correct causal result since they only measure why a model made its predictions (for the observed data)[417,418]. Empirically, this limitation has been identified for causal conclusions in longitudinal life course analysis[419].

Post-hoc model explainability metrics measure explainability after predictions have been made, and there are i) local metrics for individual predictions and ii) global metrics for the entire model[413,420]. There are many useful model-specific metrics, yet we focus in this thesis on model-agnostic metrics to ensure flexibility and generalisability of metrics to different models and settings. SHAP (SHapley Additive exPlanations) is a framework that borrows ideas from game theory – Shapley values which satisfy three axioms for game fairness. SHAP uses both permutations and surrogate models, and produces estimates of Shapley values that measure the contribution of each model covariate to a model's predictions for each observation[418]. Aggregating these local, per-prediction values in a user-defined way provides a global/overall) sense of model explainability.

From a theoretical perspective, these metrics can help build more trust and transparency in data-driven models (a key challenge for AI[413,421]). Practically, since the goal of epidemiological modelling is to understand epidemic dynamics and inform public health policy, addressing trust and transparency issues is crucial for retention of trust in AI models and for effective communication of reliable model outputs to decision makers[53,411].

Uncertainty in AI

Alongside the issues of explainability and interpretability, a modeller is also interested in the trustworthiness of the AI model and its outputs, which then determines the certainty that is expressed when communicating results to decision-makers. If an AI-based model is to be used for public health policy, understanding the extent of uncertainty is not simply a statistical issue.

Uncertainty quantification was problematic for many early machine learning and deep learning models. However, many of these models can now incorporate uncertainty in many different ways, yet at the core, there are often two underlying principles of reducible and irreducible uncertainty. We will avoid the distinction between epistemic (model) and aleatoric (data) commonly used in AI, as it lacks coherence and consistency in literature (e.g., multiple definitions and confusion between model-based predictions and external data-generating processes)[422].

Bayesian methods (e.g., Monte Carlo dropout[423]) and ensemble approaches (e.g.,[424]) have long been the pre-eminent methods to capture uncertainty in AI-based models. However, these techniques do not provide theoretical guarantees on the coverage properties of a model's prediction intervals/sets[425]. In contrast, conformal inference creates statistically rigorous prediction intervals/sets in a distribution-free way for any model that produces uncertainty in its estimates. Conformal inference relies on i) a sufficient number of independent and identically distributed (IID) pairs of training observations, ii) a calibration step involving a subset of the training data and a conformal score (to measure performance on the calibration data), and iii) a finite sample correction to the empirical quantiles[425]. For epidemiological data, the IID exchangeability assumption is immediately violated by the temporal and spatial dependencies (autocorrelation) of observations. To this end, more advanced conformal inference techniques involving control theory, while still in their infancy, have recently shown promise in producing statistically valid and sharp prediction sets for sequential data[426].

2.6.4 Interdisciplinary approaches

Throughout this chapter, we have described i) similarities between epidemiological modelling and other branches of science, ii) a wide range of influences and heterogeneous data streams to model epidemic dynamics, and iii) differing yet intersecting modelling approaches to answer similar policy/research questions.

These concepts suggest likely benefits from hybrid, interdisciplinary, and possibly multi-model approaches, as has been advocated in many reflections from past outbreaks[5,169,170,246,427,428,429]. The benefits of multiple data streams and modelling approaches can be i) the enhanced robustness of causal conclusions (e.g., synthesising evidence from different modelling assumptions[430]), ii) the improved accuracy and representativeness of modelling outputs (e.g., by incorporating additional data streams or modelling types), and/or iii) more diverse and useful information (e.g., understanding epidemic

roles of human movement and climatic conditions) for informing difficult and uncertain public health decision: the ultimate goal of analyses.

Forecasting ensembles

As we discuss in Chapters 4, 8, and 10, understanding the likely short- and medium-term future of an epidemic is often important for informing public health decision-making. For example, a forecast could promote proactive actions to mitigate the impacts of overwhelmed hospitals. Forecasting is often challenging for modellers due to aforementioned difficulties such as a possibly modifiable future and limits to predictability[3,18,103,159,161].

Recognising benefits previously observed for weather forecasting, in recent years, modellers have increasingly used interdisciplinary ensemble approaches for epidemic forecasting[431,432]. These ensemble models combine probabilistic forecasts of several different forecasting models. Across various infectious diseases, forecast targets, and combination methods, ensemble approaches have often yielded more accurate and robust probabilistic forecasts[162,171,433,434]. This is because ensemble models can combine the individual strengths of individual models, while recognising that each model is wrong, albeit in potentially different ways[103]. A model's relative strengths and weaknesses can be viewed through statistical (e.g., performance metrics) and practical (e.g., detection of incidence surges) lenses, which can be used to inform weighting of individual ensemble members for the ensemble's forecasts. For example, for the trained forecasting ensemble of the COVID-19 Forecast Hub in the US, the model weights were based on the historical performance of individual models, as measured by the Weighted Interval Score (WIS)[58,172].

Causality in epidemics

Forecasting ensembles are just one example of the benefits of reaching beyond a single modelling approach. A modeller can also gain quantitative understanding by using techniques from different domains separately.

The infeasibility of performing real-world experiments limits a modeller's ability to understand and quantify the *causes* of an epidemic – a complex, stochastic, and highly nonlinear dynamical system with feedback loops[99]). For example, cluster-randomised controlled trials during an epidemic would likely be unethical, despite possibly providing valuable policy insights into the effectiveness of different public health interventions[90,435]. These resource-intensive trials would also likely take a long time to scale up, and therefore lack the timeliness required for fast-evolving epidemic dynamics. This creates a distinction between causal inference for the study of epidemics and for traditional epidemiology, where case-control, cohort, and clinical trials are all feasible and useful in the latter setting.

However, modellers are aided by the increasing availability of i) data, potentially from other streams[19] such as wastewater[436] or behaviour/social[234] data, ii) causal inference methods such as structural causal models[437], and iii) other mathematical and statistical tools beyond traditional modelling approaches such as wavelet methods and block cross-validation[79,438]. All of these present promising avenues for understanding why, when and where an epidemic is evolving or has evolved[439,440]. Recent examples of their usage in infectious disease modelling include[79,89,438,441].

Physics, statistics, deep learning, and weather forecasting

In the weather forecasting and climate modelling communities, it is increasingly common to combine physics and statistics with machine learning and deep learning for solving and simulating the complex dynamical systems of weather and climate[287]. Epidemic systems are not subject to the same carefully-derived governing law of physics[288], yet there are similarities as modellers can: i) embed physical mechanisms into the deep learning model (e.g., a PINN[400]) of an epidemic via the model architecture/structure or custom loss functions, ii) use neural networks (e.g., deep operator networks or Fourier neural networks) to solve a system of ODEs or PDEs, and iii) blend statistical principles (e.g., Bayesian inference)

with deep learning architectures to achieve latent representations for dynamical systems where modellers can more easily simulate the system[287,289].

As discussed in a recent review of 245 studies that integrated AI with mechanistic epidemiological modelling[289], such approaches are well-suited to addressing many of the complexities posed by modelling infectious diseases. Although these approaches are in their infancy[289], they have recently demonstrated their potential for inferring epidemic drivers and dynamics for various infectious diseases[384,400,442].

Importance of interdisciplinary modelling

"When a law is right it can be used to find another one. If we have confidence in a law, then if something appears to be wrong it can suggest to us another phenomenon"[2]. As featured in the thesis prologue, the excerpt from Feynman[2] is representative of the thesis' ambition and consistent preference for scientific discovery from different disciplines: different modelling approaches, data streams, and branches of science (Figures 1.1, 2.2).

The idea aligns with the recent movement towards scientific understanding during epidemics from consensus approaches that rely on multiple models and multiple disciplines[18,172,428,443]. Glancing at historical literature beyond epidemiological modelling, scientific discoveries, from the ancients to Copernicus, have often been made independently, as alternative assumptions were tested to enhance the outputs of different analyses and derive more general laws. In anticipation of future chapters, the contributions of this thesis (Table 1.1) will not be compared with Newton's Laws nor the generalisation from the Law of Gravitation. However, we will embrace similar logic of method diversity when working to enhance scientific discovery and improve public health decision-making.

2.6.5 Simulation, inference and optimisation

When simulating a model or fitting a model to data (i.e., inference), the given knowledge (and available data for inference) and the research question for the infectious disease generally inform the modelling approach and model structure/architecture[188,444]. This in turn guides the choice of simulation, inference, or optimisation algorithm which is the vehicle to implement the chosen model, potentially using the available data for inference[15,349]. The optimisation algorithm is a similar concept for machine learning and deep learning models. The choice of algorithm is a fundamental engineering/design choice that can influence both the accuracy and efficiency of the modelling[15,445,446]. Epidemic data are incomplete, imperfect and potentially high-dimensional (Section 2.4), and model outputs can often be needed rapidly to inform decision-making about fast-evolving epidemic dynamics[17,20,165]. All of this highlights the importance of accuracy and efficiency considerations.

Algorithms for mechanistic models

Mechanistic modellers historically used either analytical or numerical techniques for simulation and model fitting. Analytical solutions to mechanistic models are rare and mainly occur for simple models without nonlinearity issues. This has motivated a range of numerical integration schemes. For systems of differential equations, there are both forward (solving for a given set of parameter values) and inverse (inferring parameter values from data) problems. The simplest explicit approach to the forward problem is Euler's method of integration[122], a first-order numerical procedure for solving systems of ODEs with specified initial values. It can be extended to solving more complex systems of SDEs (Euler–Maruyama method) and PDEs (finite differences). The Euler method is the simplest of the Runge-Kutta methods – a family of higher-order, iterative methods that can enable greater accuracy and stability for more complex dynamical systems.

For simulating stochastic models, exact event-based approaches include the classical Gillespie direct and first reaction simulation algorithms[50,51,349,447]. Designed by a physicist called Daniel T. Gillespie, Gillespie's direct algorithm works by

jointly determining the time and type of event (using the propensity at which total events occur), whereas the first reaction algorithm determines the time and type of event based on the minimum of individual event propensities[50,51]. Central to the choice of a numerical simulation algorithm and related hyperparameters are setting-specific considerations of stability, accuracy, and computational resources. We will discuss and extend these algorithms in detail in Chapter 7.

Before discussing inference routines for mechanistic models, we note that simulation and inference are not stand-alone activities. Solving the forward problem, and using simulation schemes more generally, can be used for solving the inverse problem (i.e. for inference). Simulation routines are also essential for verifying and improving the accuracy and efficiency of inference algorithms[188].

More generally, it is most often statistics alongside data that allow a modeller to answer public health policy/research questions when using a mechanistic model[122,448], as the model is paired with available data in a statistically rigorous way. In doing so, inference unlocks the potential of the models[108]. Historically, classical likelihood-based methods such as maximum likelihood were used for mechanistic-model-based inference of key epidemic parameters[62,122,448,449]. However, such methods can struggle in real-world epidemic settings for increasingly complex data, as modellers need to incorporate information and uncertainties from more types and greater quantities of data[445,446].

Theoretical advances in sampling algorithms and improvements in computational resources and software have enabled Bayesian approaches to inference[109,188,450,451]. Conditional on the availability of a likelihood for the observed data, a modeller can use Markov Chain Monte Carlo (MCMC) methods such as Gibbs sampling[452] or Hamiltonian Monte Carlo (HMC)[453]. These methods can then be paired with a simulator to solve the inverse problem. Combining the likelihood with specified prior distributions and suitable data in an MCMC algorithm allows modellers to estimate potentially intractable posterior distributions.

The choice of MCMC algorithm is non-trivial. Increased model complexity can yield highly inefficient and infeasible exploration of the target posterior distribution for classical MCMC methods, for which HMC[445] or ABC[446] may help. The data themselves can cause many further challenges (see Section 2.4). As true epidemic states are never observed[108,109], to jointly model both the observation process (e.g., observed new disease cases) and dynamic process (e.g., new infections) over time, particle Markov-chain Monte Carlo, PMCMC, can improve upon standard MCMC methods. PMCMC targets inference of Hidden Markov Processes and works by embedding Sequential Monte Carlo (SMC) within a standard MCMC algorithm such as Metropolis-Hastings[109,450,454]. SMC/particle filter methods provide estimates for the posterior distribution of the dynamic process (i.e., hidden states) at each time step, conditional on unknown parameters. The methods work by using iteratively updated and likelihood-weighted samples/particles over time[15,109,454,455]. In doing so, PMCMC exploits the temporal structure of the data. PMCMC can handle high-dimensional systems, for example using block particle filter extensions for spatially structured epidemic models[456]. However, high-dimensional systems can induce high computational costs[15,454]. If the likelihood of the observed data for a given model is intractable or unavailable, ABC may again be useful[446,457], but ABC also suffers from computational barriers and is more suitable for systems with relatively few parameters.

Algorithms for statistical models

For statistical models such as Bayesian hierarchical models, these Bayesian inference algorithms have also proved useful. HMC, often implemented in the Stan probabilistic programming language[458], provides efficient exploration of the parameter space by exploiting gradient information[188,445].

However, for spatiotemporal modelling tasks such as disease mapping across space and time, high dimensions can again impose computational burden on MCMC algorithms[15,454,459,460,461]. This motivates Integrated Nested Laplace Approximation (INLA), an approximate and computationally efficient Bayesian inference method for latent Gaussian models (a wide class of models). INLA uses nested formulations to implement analytical and numerical approximations of

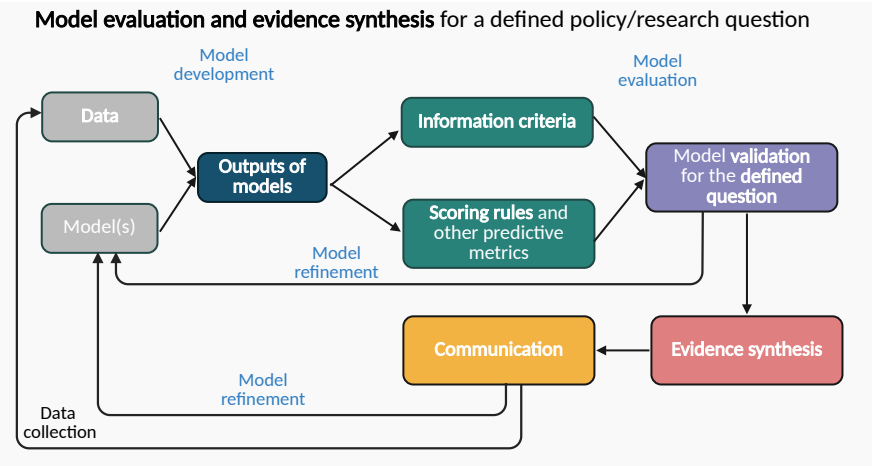


Figure 2.4: **Model evaluation and evidence synthesis:** We visualise how model evaluation and evidence synthesis fit into the overall modelling process. Starting from data and one or more models, model evaluation often involves information and/or predictive criteria. The goal of using these criteria for model evaluation is to validate a model or set of models for a defined policy/research question within a specified context, and generally involves sensitivity analyses and multiple model comparison to achieve model refinement. Conditional on having validated the model, evidence synthesis and communication are the necessary next steps to advance scientific understanding and public health policy. In real-world settings beyond this thesis, these can then guide further model refinements and data collection. The blue labels in this visualisation represent the three steps of the high-level modelling procedure presented in Figure 2.2 C.

sparse matrices. This results in fast but approximate Bayesian inference for posterior distributions of the latent field, and avoids potential convergence issues of MCMC methods[237,459].

Algorithms for AI models

Due to their data-driven nature, AI-based models are trained using optimisers. These algorithms adjust model weights/parameters to minimise a user-specified loss or error function[398]. The loss function measures the discrepancy between the observed data and the model's predictions.

As numerical routines are highly inefficient, commonly used optimisers for AI-based models include gradient descent (the basic optimiser) and stochastic gradient descent (an accelerated version of gradient descent)[398]. Both optimisers use the gradients of the loss function to iteratively update the model's parameters[462]. Gradients are computed using backpropagation at every epoch (or training iteration), and parameters are updated in the direction of the gradient of steepest descent. The stochastic approximation uses an estimate of the actual gradient. The Adaptive Moment Estimation (Adam) optimiser is a widely used optimiser for deep learning models that extends stochastic gradient descent by adding adaptive learning rates and momentum to handle sparse gradients[463].

2.7 Model evaluation and evidence synthesis

Before discussing theory for evaluating models and synthesising scientific evidence, we again step back from the modeller's perspective. A modeller generally has incomplete knowledge of the past, present and future epidemic dynamics, and their task is to answer (often rapidly), via an epidemiological model, specific policy/research questions[17,103,160,203]. As noted by Medley [18], without such defined questions to answer, the model outputs are just lines on graphs.

Conditional on fitting the model to observed data or simulating different epidemic trajectories, the utility of the model is most often determined by its ability to reliably inform decision-makers in a timely manner about realistic current and future public health policy options[16,18,108,160]. Therefore, validating the utility of a model for the defined explanatory or

predictive task is essential and can guard against the erosion of public and decision-makers' trust.

2.7.1 Model evaluation: Information criteria and scoring rules

Model evaluation or validation (Figure 2.4) is a component of model comparison for the given policy/research question. Evaluation is often defined as the determination of whether a model is sufficient and appropriate, within a specific context[103,241], to explain the observed data[62]. The context is often the defined policy/research question, the temporal and spatial scales, and the local decision-maker in question, while the observed data might be reported disease cases or hospitalisations. Both the data and model are only "good" conditional on each other and the overall context[103,105,199]). Model evaluation therefore depends on both the context-specific quality of the model and available data. Evaluation procedures rely on formal statistical analyses involving predictive criteria such as information criteria and scoring rules[62].

Information criteria are predictive measures that generally emphasise model parsimony in model selection; balancing likelihood (fit to data) against model complexity (number of estimated parameters). Applicable to both mechanistic and statistical models, classical examples include the Akaike Information Criterion (AIC)[464] and Bayesian Information Criterion (BIC)[465]. Both AIC and BIC make potentially strong assumptions. These include the lack of uncertainty in parameter estimates for AIC, or assumptions of a true model and no measurement of out-of-sample performance for BIC. More recently, leave-one-out (LOO) estimates of the Expected Log pointwise Predictive Density (ELPD) have been proposed – ELPD-LOO. These are computed using LOO cross-validation (LOO-CV) and estimate the model's out-of-sample generalisability by computing the predictive distribution for new unseen data[391,466]. ELPD-LOO estimates are more robust measures for model comparison, and have more recently been used in infectious disease modelling. For example, Hozé et al.[319] compare different serocatalytic models using ELPD-LOO. However, care is again needed for cross-validation when using the spatially and temporally structured data of epidemics[391,439].

For deep learning models, a penalty for the number of parameters is irrelevant as these models are, by construction, complex (potentially millions of parameters), and do not have an explicit likelihood function, which means a different, often predictive approach is required.

Historically, when measuring the quality of predictions for an epidemiological model, pointwise accuracy metrics were used, including the Mean Absolute Error (MAE), R^2, or Root Mean Squared Error (RMSE)[467,468]. These accuracy metrics are applicable to mechanistic, statistical, and deep learning models, but for reliable policy guidance, we are also interested in the ability of a model to correctly quantify uncertainty in its forecasts[159].

Building on advances in economic and weather forecasting, so-called proper scoring rules have been adopted that measure the whole predictive/forecast distribution. Since the COVID-19 pandemic, these metrics have become commonplace in forecasting and nowcasting studies[167,170,173,433,469]. For example, the Continuous Ranked Probability Score (CRPS) generalises the MAE for probabilistic forecasts and simultaneously measures calibration and sharpness[470]. CRPS is therefore a measure of distributional accuracy and is defined for a forecast's cumulative distribution function (CDF) F and observation y as:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - 1(x \geq y))^2 dx. \tag{2.7.1}$$

Some models only permit forecasts using quantiles of the CDF F. Examples include deep learning models with quantile regression. Quantiles are also computationally less intensive to process than samples from the predictive distribution. Therefore, forecasts are often represented using quantiles[167,433,434], for which the recently developed WIS can be used as an approximation to the CRPS[58]. These scoring rules were designed for forecasting, but can also be used to measure the distributional accuracy of parameter inference[141]. We will explore scoring rules and their assumptions in further detail in Chapter 8.

2.7.2 Multiple model comparison, evidence synthesis, and communication

In the above, we focus on model evaluation. Model evaluation is a step along the iterative process of model development, evaluation, and refinement (see Figure 2.3 and [188]). For infectious disease modelling, these processes are not a purely statistical/theoretical exercise. Although the role of scientists and modellers is not to advocate or translate evidence into policy, it is necessary to develop scientific understanding and synthesise evidence to inform decisions made by public health authorities [18]. The scientific understanding may come from one or more models, and each model produces their "best" possible estimate of key parameters or future epidemic trajectories (Table 2.1). Then, there are public health, societal, and economic consequences for any subsequent policy decisions. Therefore, model evaluation and iterative model refinement (Figure 2.2 and [188]) focus on the model's ability to infer and/or predict, inclusive of uncertainty, quantities that are relevant for difficult local policy decisions. Examples include inferring $R(t)$ for COVID-19 [107,134,454] or forecasts for dengue [159,162].

In terms of evidence synthesis, there are many formal techniques, ranging from systematic reviews to meta-analyses. A recent trend of using multiple models (see Section 2.6.4) often involves different researchers, assumptions, and data streams to answer the same policy/research question [363,443,471,472]. Although the public health utility of the models' outputs will still be bounded above by the quality of individual models and the weighting assigned to evidence from individual models [433,443,471], the concept of developing understanding from multiple models has been used increasingly for forecasting [172,433,434,473], scenario planning [18,443], and intervention modelling [472].

These empirical findings of utility are complemented by theoretical grounding in decision theory [160] and causal inference [430] (e.g., triangulation), and again reflect the fact that all models are inherently wrong and uncertain, yet they can be useful in different ways for different purposes in different settings [103,160,199].

In real-world epidemic settings, the identified abilities and limitations of a model, including what a model cannot tell us, can then be communicated to the relevant stakeholders along the decision-making process (Figure 2.2 D) [103,105,160]. The statistical theory and modelling details are essential for model comparison. Meanwhile, the step from evidence synthesis to dissemination prioritises succinct, clear, and prompt communication of outputs to decision-makers and other stakeholders (e.g., scientific/policy advisors). These procedures have been documented i) in recent guidelines for digital prediction tools [280] and pandemic modelling [20,160], and ii) empirically by collaborative pandemic modelling efforts such as the SPI-M-O committee and Scenario Modelling Hub during the COVID-19 epidemic in the UK and US respectively [18,443].

Communication of scientific output is non-trivial, clarity of audience is crucial, and is ideally paired with decision theory, social science, behavioural studies, and other setting-specific considerations [3,105,160]. All such considerations are discussed in our forecast evaluation manuscript and in the discussion of this thesis' contributions (Chapters 8 and 10 respectively).

Chapter 3

The utility of wastewater surveillance for monitoring SARS-CoV-2 prevalence

Cathal Mills, Marc Chadeau-Hyam, Paul Elliott, and Christl A Donnelly. The utility of wastewater surveillance for monitoring SARS-CoV-2 prevalence. *PNAS Nexus*, page 438, October 2024. ISSN 2752-6542. doi: 10.1093/pnasnexus/pgae438. URL https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae438/7811170

We performed the first nationwide modelling analysis for the utility of wastewater surveillance for monitoring SARS-Cov-2 community prevalence in high-spatial resolution across 21 months of the COVID-19 pandemic. The manuscript was published in *PNAS Nexus* in October 2024.

Abstract

Public health authorities have increasingly used wastewater-based epidemiology (WBE) to monitor community transmission of SARS-CoV-2 and other agents. In this study, we evaluate the utility of WBE during the COVID-19 pandemic in England for estimating SARS-CoV-2 prevalence. We use wastewater data from the Environmental Monitoring for Health Protection program and prevalence data from the REal-time Assessment of Community Transmission-1 study. Across the pandemic, we describe how wastewater-based modelling can achieve representative SARS-CoV-2 prevalence estimates in fine and coarse spatial resolutions for relatively short-time horizons (of up to 1 month), and thus assist in filling temporal gaps in surveillance. We infer a temporally evolving relationship between wastewater and prevalence which may limit the utility of WBE for estimating SARS-CoV-2 prevalence over longer time horizons without a concurrent prevalence survey. Exploring further our finding of time-varying, population-level faecal shedding, we characterize WBE for SARS-CoV-2 prevalence as (i) vaccination coverage dependent and (ii) variant-specific. Our research suggests that these factors are important considerations in future uses of WBE by public health authorities in infectious disease outbreaks. We further demonstrate that WBE can improve both the cost efficiency and accuracy of community prevalence surveys which on their own may have incomplete geographic coverage and/or small sample sizes. Therefore, in England, for the objective of high spatial resolution prevalence monitoring, strategic use of SARS-CoV-2 wastewater concentration data nationally could have enhanced, but not replaced, community prevalence survey programs.

Author Contributions

- Conceptualisation: **C.M**; C.A.D
- Methodology: **C.M**; C.A.D.; P.E; M.C.H
- Data curation and visualisation: **C.M**
- Writing original draft: **C.M**; C.A.D
- Writing final draft: **C.M**; C.A.D; P.E; M.C.H

3.1 Introduction

Monitoring the distribution of viruses has been a common use of wastewater-based epidemiology (WBE) in recent years[292]. WBE involves collection of samples from wastewater sites that capture human excretions through urine or faecal matter, and has enabled, for example, surveillance of polioviruses[291]. During the COVID-19 pandemic, as governments and public health authorities around the world sought to monitor the transmission and spread of SARS-CoV-2 (the virus which causes COVID-19), WBE was applied as a infection surveillance tool, since SARS-CoV-2 is excreted in the faeces of both symptomatic and asymptomatic infected individuals.

Clinically confirmed SARS-CoV-2 prevalence estimates are sensitive to test-seeking biases, asymptomatic infections, and clinical testing capacity. Conversely, WBE represents an indirect, non-invasive, population-level infection surveillance tool for cost-effective, real-time monitoring of pathogen transmission[474,475]. Several studies have established strong correlations between wastewater SARS-CoV-2 concentrations and reported COVID-19 cases[294,295,476], whilst others estimated the time dependency and lead time (varying from four to ten days) from detection in wastewater to date of testing in clinical cases[293,477,478]. During the pandemic, SARS-CoV-2 prevalence in England was estimated for 45 sewage site catchments[479], whilst further research estimated weekly viral wastewater concentrations in a spatially continuous domain[480]. More recently, the issue of how best to incorporate wastewater surveillance into a prevalence estimation framework has been explored with a specific focus on nowcasting community prevalence in counterfactual scenarios[481].

Here, across 21 months of the COVID-19 pandemic in England, we provide a high-resolution spatiotemporal analysis and evidence synthesis of the utility of WBE for estimating community SARS-CoV-2 prevalence. Our analysis uses data from one of the world's largest community prevalence surveys; the REal-time Assessment of Community Transmission-1 (REACT-1) study, and the Environmental Monitoring for Health Protection (EMHP) wastewater surveillance programme. Our geospatial framework maps wastewater concentrations from a sewage treatment plant to the level of a Lower Tier Local Authority (LTLA). Then, our analysis provides a rigorous investigation of the relationship between wastewater concentration estimates and estimated infection prevalence from REACT-1. We investigate the stationarity of this relationship over time and whether we can disentangle factors which influence the prevalence-to-wastewater relationship. Leaning on this exploratory analysis, we perform extensive modelling analyses to investigate the spatial and temporal extent to which WBE can facilitate reliable estimation of community prevalence of SARS-CoV-2. Our analyses and the lessons learned therein will serve as recommendations for future infectious disease outbreaks where authorities wish to combine infection surveillance by survey data and WBE.

3.2 Results

Our analysis, across two distinct studied periods of the COVID-19 pandemic (labelled *early* and *late* periods which are segmented due to surveillance differences - see Materials and methods), focuses on how SARS-CoV-2 wastewater

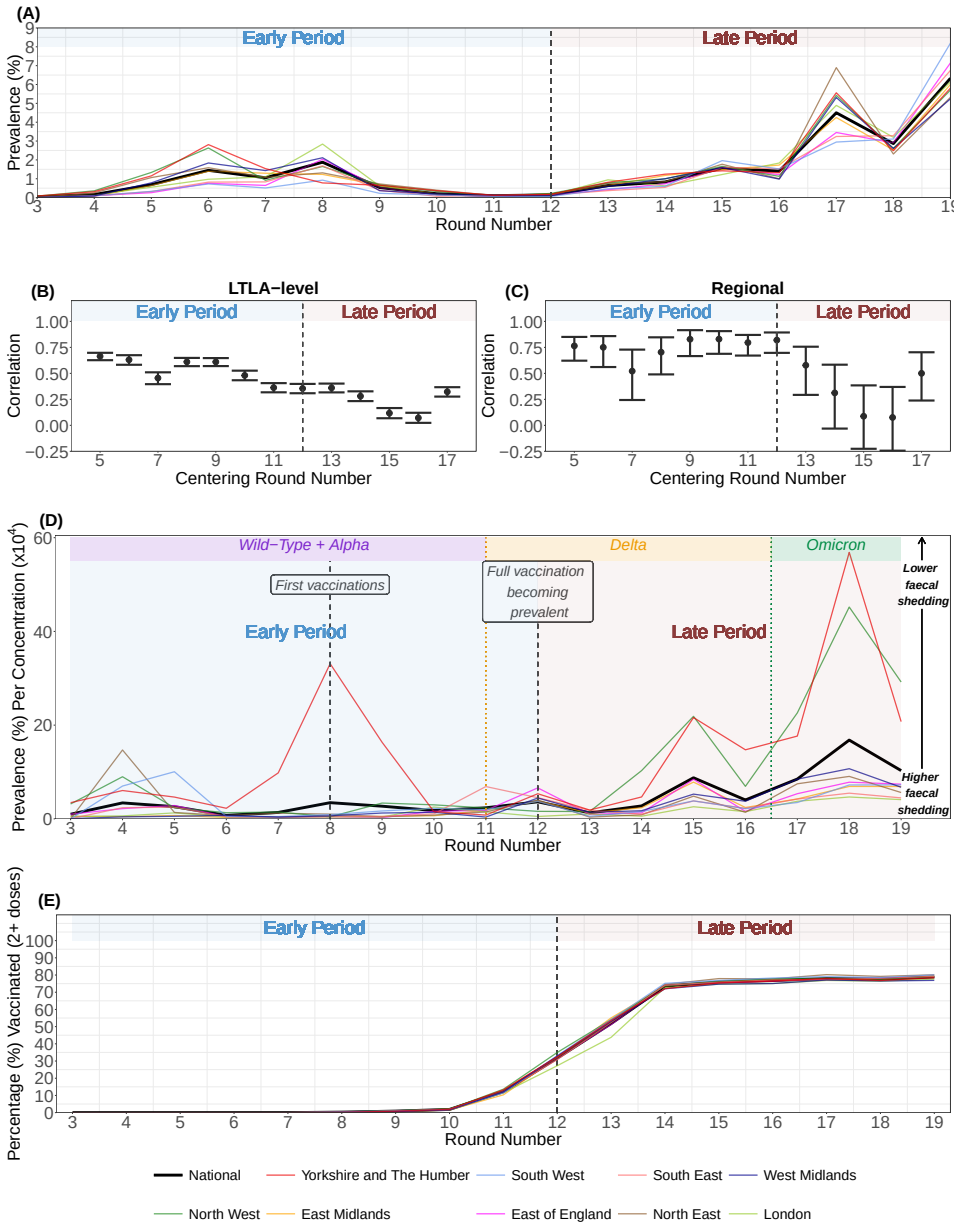


Figure 3.1: **Prevalence-wastewater relationship.** A) Population-weighted regional (and national) averages of SARS-CoV-2 prevalence estimates per survey round of the REACT-1 study. B) Spearman's correlation between all LTLA-level observations of SARS-CoV-2 wastewater concentrations and REACT-1 SARS-CoV-2 prevalence across moving five-round windows centred on each survey round from REACT-1 rounds 5 (commencing 18 September 2020) to 17 (ending 20 January 2022). 95% Confidence Intervals (CIs) were obtained by the Clopper-Pearson method[482]. The number of observations used for computing the correlations across each five-round window varied depending on the number of LTLAs covered by the wastewater surveillance programme. C) Analogous visualisation of correlation between all observations of regional wastewater concentration and SARS-CoV-2 prevalence estimates within the same moving five-round windows ($n = 45$ for each window). Regional estimates were derived by population-level weighting of the corresponding LTLA-level wastewater concentrations and prevalence estimates. To visualise how any LTLA-and-region-specific heterogeneities in wastewater concentrations and/or prevalence estimates impact their individual prevalence-wastewater relationships, we also calculated correlations within each LTLA (region) separately (visualised in Figure SI A.1). D) Population-weighted regional (and national) REACT-1 prevalence per population-weighted regional (and national) estimated wastewater concentration by round of data collection, from REACT-1 rounds 3 (24 July 2020) to 19 (31 March 2022). Greater values of the ratio correspond to lower implied population-level faecal shedding rates i.e. per positive infected individual. For reference, vaccination began in round 8 (6 January 2021) with full vaccination becoming prevalent from round 12 (20 May 2021) onwards (when the survey round estimate of the national average proportion of the entire population fully vaccinated is 31.7%), the Delta variant was dominant from rounds 11-16 (15 April 2021 - 14 December 2021), and the Omicron BA.1 and BA.2 sub-variants are dominant in rounds 17 to 19 (from 5 January 2022 to 31 March 2022). E) Population-weighted regional (and national) proportions of the population with two or more doses of a licensed vaccine, as estimated per survey round of the REACT-1 study. Note that only data from REACT-1 rounds 3 to 19 are displayed as this represents the time period which overlaps with the wastewater surveillance programme.

concentration estimates can be associated with the community SARS-CoV-2 prevalence estimates from individual rounds of the REACT-1 survey. We consider varying spatial scales in our analyses; from coarser (regional) to fine (LTLA) resolution. Our wastewater-based, gradient boosting model learned the relationship between wastewater concentrations and survey prevalence estimates during model training. The wastewater-based model was then used to estimate out-of-sample (i.e. hypothetically unobserved) SARS-CoV-2 prevalence within individual prevalence survey rounds. For context, the aforementioned *early* period of analysis (rounds 3 to 11, from 24 July 2020 to 3 May 2021) was the first period of overlap between the wastewater surveillance programme and REACT-1 survey. The period corresponded to a duration of relatively low, stable SARS-CoV-2 prevalence levels (mean LTLA estimate: 0.55%) and vaccination was at low levels (mean fully vaccinated proportion for an LTLA in round 11 was 12.2%). Conversely, the ten-month *late* period (rounds 12-19, 20 May 2021 to 31 March 2022) incorporated expanded wastewater testing coverage and the epidemic's evolution was characterised by high prevalence levels, heterogeneous vaccination of the population, and occurrence of a new, highly transmissible Omicron variant with different properties, such as lower viral loads, selective reduction in Omicron infectivity in the intestinal tract, and shorter duration of respiratory shedding, compared to the Delta variant[483,484,485].

3.2.1 Relationship Between SARS-CoV-2 RNA Wastewater Concentrations and Prevalence

Before assessing how a wastewater-based model may be used to estimate SARS-CoV-2 prevalence, we describe the relationship between SARS-CoV-2 prevalence estimates and corresponding wastewater concentrations, at varying time periods throughout the COVID-19 pandemic. Consistent with the underlying scientific premise, in the *early* period (REACT-1 rounds 3 to 11, from 24 July 2020 to 3 May 2021), estimated wastewater concentrations at an LTLA level were moderately to-strongly correlated with SARS-CoV-2 prevalence (Spearman's correlation $r = 0.62$; 95% CI: $0.59, 0.65$). These trends were observed at fine spatial resolution (LTLA), and stronger relationships were observed at coarser spatial resolution between (population-weighted) regional and national averages wastewater concentration and corresponding (population-weighted) averages of prevalence estimates (Table SI A.1). Looking at the time evolution of the prevalence-wastewater relationships (Figure 3.1), as measured by correlations between all wastewater concentrations and SARS-CoV-2 prevalence within moving (centrally aligned) windows of five survey rounds, we estimated relatively stable, consistently strong relationships until the emergence of the Delta variant and the onset of increasingly widespread vaccination in the population (from rounds 11 and 12 onwards - which began on 15 April 2021 and 20 May 2021 respectively). Indeed, across the entire *late* studied period (rounds 12 to 19, from 20 May 2021 to 31 March 2022), we found the relationship between concentrations and prevalence to be more complex, volatile, and temporally evolving (Figure 3.1). For example, within a subset of five pre-Omicron rounds (rounds 12 to 16, 20 May 2021 to 14 December 2021) of the *late* period, we report weaker correlation ($r = 0.28$; 95% CI: $0.23, 0.33$) between LTLA-level concentrations and prevalence. Similarly diminished relationships were observed between the regional population-weighted prevalence estimates and corresponding wastewater concentrations, despite the increased geographic coverage of the wastewater surveillance programme during this period. We note that these trends of diminishing correspondence were similarly observed if we assessed the correlations (over time) between prevalence and concentration estimates within each LTLA (or region) separately (Figure SI A.1).

To further investigate these findings of time-varying complexities, we propose here a new metric (which may be a useful tool for future infectious disease outbreaks) - the estimated prevalence per concentration in wastewater (Figure 3.1 D). The metric captures the prevalence-to-wastewater relationship, and so population-level faecal shedding over time (with higher values indicating lower average faecal shedding per person). Relationship complexity (suggested previously by the aforementioned five-round-windowed correlations and now in the prevalence per concentration) was evident in the presence of i) rapid rollout of vaccination of LTLA populations nationally, and ii) rapid replacement of the Delta variant by the Omicron BA.1 and BA.2 sub-variants (Figure 3.1). In terms of the impacts of vaccination on the prevalence-to-wastewater relationship, we observed substantially diminishing correlation between prevalence estimates and wastewater

concentrations (Figure 3.1) during five-round windows where full vaccination was increasingly prevalent (i.e. from rounds 12 onwards). In terms of our proposed metric, across all observations from rounds 12 to 19 (from 20 May 2021 to 31 March 2022), the estimated prevalence per wastewater concentration was moderately correlated ($r = 0.45$; 95% CI: $0.41, 0.48$) with the corresponding (estimated) proportion of an LTLA population that was fully vaccinated (two or more doses of any vaccine), though correlation was stronger when we instead considered the population proportion fully vaccinated in the previous round ($r = 0.51$; 95% CI: $0.48, 0.54$). Similarly, focusing on the evolution of vaccination-prevalence per concentration over time within LTLAs, the median of the LTLAs' correlations ($n = 309$), calculated using each LTLA's observations across eight survey rounds, between prevalence per concentration and fully vaccinated proportion from the previous round was 0.64.

Thus, in understanding the association between SARS-CoV-2 prevalence and wastewater concentrations, it is important to account for the proportion of a population fully vaccinated and specifically the percentage of vaccinated individuals in the preceding month. The implied lag may reflect time taken for a vaccine to become effective at reducing faecal shedding. Inclusion of the vaccination-concentration interaction (see Table SI A.2) increased the wastewater-infection prevalence correlation ($r = 0.71$, 95% CI: $0.69, 0.73$) relative to the unadjusted model.

By mid-December 2021, the REACT-1 study and EMHP wastewater programme had found that Omicron BA.1 had become the dominant SARS-CoV-2 variant in England[302,486]. Here, we show that the percentage testing positive per concentration in wastewater increased during the Omicron period, indicating less faecal shedding per infected individual at a population level (Figure 3.1). Reduced levels of detected population-level faecal shedding were observed as wastewater concentrations did not show the same magnitude of increases as prevalence levels. Such reduced population-level shedding was observed during survey rounds and there was initially a breakdown in the longitudinal correspondence between LTLA-level concentration and prevalence (around REACT-1 round 17), although the correlation strengthens when we narrow our focus to this new period with widespread, reduced population-level faecal shedding (Figures 3.1 and SI A.1).

3.2.2 Out-of-Sample Wastewater-Model-Based Estimates of REACT-1 Prevalence

Moving towards our wastewater-based modelling analysis, we now describe results in out-of-sample environments, where estimates were out-of-sample in the sense that we omitted the prevalence data for each (entire) individual testing survey round (and all subsequent rounds). To assess the predictive performance of our (wastewater-based) model's in estimating the prevalence, our out-of-sample approach split the data into training and testing sets. We considered two data splitting strategies: i) iterative estimates where we estimated community SARS-CoV-2 prevalence within a single survey (testing) round using a wastewater-based model trained/calibrated on all prevalence data up to one round before each testing round, and ii) multi-step estimates where we estimated prevalence for three individual (testing) survey rounds using a wastewater-based model where the model is not recalibrated across the three testing rounds."

Employing an iteratively updating wastewater-based, gradient boosting model, in the *early* period, from REACT-1 rounds 7 to 11 (3 November 2020 to 3 May 2021), we report weak-to-moderate within-round predictive accuracy for estimating LTLA-level REACT-1 prevalence, which is reflected in Mean Absolute Error (MAE) values, occasional failure to detect which LTLAs were experiencing the most extreme prevalence levels across the nation, and sub-optimal accuracy for detection of prevalence level changes (Table SI A.3). These results are reflective of modest predictive abilities of the wastewater-based model, yet also collectively suggest occasional difficulties in relying exclusively on a wastewater model (post-calibration) for high-resolution prevalence inferences. Nevertheless, if we assess the iteratively updating wastewater-based model's out-of-sample estimates collectively (i.e. over time) across the five survey rounds, there was moderate-to-strong predictive accuracy (MAE = 0.37%) and strong correspondence between the model's estimates and the survey prevalence estimates, both within LTLAs' individual time series (median $r = 0.85$) and across all LTLA-level observations ($r = 0.66$). Moving towards a coarser regional geographic scale (where population-weighted regional mean

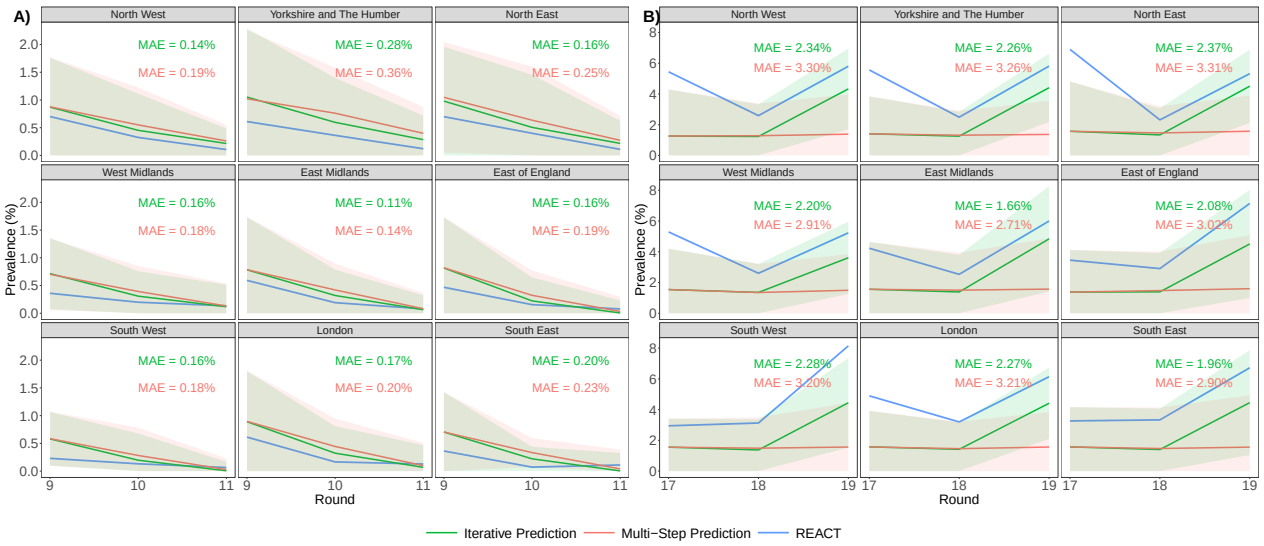


Figure 3.2: **Regional out-of-sample, wastewater-model-based prevalence estimates in A) rounds 9 to 11 (4 February 2021 to 3 May 2021) and B) rounds 17 to 19 (5 January 2022 to 31 March 2022).** Out-of-sample wastewater-model-based prevalence estimates are shown alongside REACT-1 prevalence (blue), where estimates were out-of-sample in the sense that we omitted the prevalence data for each (entire) individual testing survey round (and all subsequent rounds). The wastewater-based model was trained at an LTLA level i) with iterative (green) updating of the model (thus expanding the training set every round with an additional training round), and ii) without model recalibration (red - which means there was a fixed training set size across the three testing rounds). The two scenarios amount to employing a wastewater-based model for either a single survey round (i.e. a relatively short period of time up to one month) or for three survey rounds (i.e. a more sustained period of up to three months), and thus enable comparison of predictive performance with, and without, regular model calibration of the wastewater-to-prevalence relationship. The wastewater-model-based and REACT-1 prevalence estimates for each region (per individual round) were obtained by weighting the LTLA-level prevalence estimates by the corresponding LTLA populations within the region. The shaded 95% prediction intervals (PIs) were obtained by 5,000 non-parametric bootstrap samples. In A), the mean regional (population-weighted) average prevalence across rounds 9 to 11 was 0.28%, and in B), the mean regional (population-weighted) average prevalence across rounds 17 to 19 was 4.57%. The substantial differences in prevalence levels are the reason for the different y-axes for prevalence across A) and B).

prevalence estimates were obtained by weighting the LTLA-level estimates by their corresponding populations), there was greater representativeness in the wastewater-based model's estimates (Figures 3.2 and SI A.3, Tables SI A.3 – SI A.4). The superior predictive accuracy is reflected in the ability to reliably estimate regional average prevalence (MAE = 0.27%) and the accuracy of 88.9% (95% CI: 75.9%, 96.3%) in estimating whether the regions were experiencing increasing or decreasing prevalence levels.

Remaining within the *early* period, but focusing on REACT-1 rounds 9 to 11 (4 February 2021 to 3 May 2021) to enable comparison between the iteratively updated and multi-step estimates produced for these three rounds, we observed relatively similar, weak-to-moderate predictive accuracy at an LTLA level for the iteratively updating and multi-step, wastewater-based model's estimates (Table SI A.5). At a regional level, both the iteratively updating and multi-step wastewater-based model's estimates were strong at estimating prevalence movements (i.e. increasing or decreasing prevalence levels) and the correlation between the model-based estimates and REACT-1 survey-based prevalence was strong in both settings. Nevertheless, the iteratively updating model's (population-weighted) regional prevalence estimates benefited from the recalibrations of the prevalence-to-wastewater relationship as its estimates (n = 27) outperformed across the metrics of MAE (Iterative: 0.17%, Multi-step: 0.21%), detecting prevalence movements (Iterative: 92.6%, Multi-step: 77.8%), and correlations with the prevalence (Iterative: 0.94, Multi-step: 0.93) (Table SI A.5).

Moving towards the *late* period (20 May 2021 to 31 March 2022) with higher prevalence levels and increasing population vaccination levels, the iteratively updating wastewater-based model's out-of-sample prevalence estimates, at an LTLA level for REACT-1 rounds 15-19 (19 October 2021 to 31 March 2022), indicated moderate predictive power in terms of detecting

prevalence movements (75.4%, 95% CI: 73.2%, 77.5%) and tracking precise prevalence levels (r = 0.65) (Tables SI A.6 –SI A.7). Similarly, looking at the out-of-sample estimates for the five rounds collectively (i.e. over time), there was once again a moderate-to-strong correspondence between our LTLA-level model-based estimates and survey-based prevalence (r = 0.65, MAE= 1.65%) and likewise, when analysing the correspondence within individual LTLAs separately (Median r = 0.78, Q1= 0.57, Q3= 0.90). Largely similar trends were observed in coarser regional spatial resolution when we weighted our LTLA-level (wastewater-model-based prevalence) estimates by their corresponding populations within the regions (Figure SI A.4, Tables SI A.6 – SI A.7).

Nevertheless, despite the iteratively updating model's recalibration of the prevalence-to-wastewater relationship every round (i.e. an expanding model training window), the onset of the Omicron BA.1 peak in REACT-1 round 17 (5 January 2022 to 20 January 2022) was largely underestimated and undetected by our wastewater-based model (Table SI A.6, Figure 3.2). The LTLA-level prevalence estimates departed from survey-based prevalence levels (MAE = 2.76% and movement detection accuracy of 62.8%, 95% CI: 57.1%, 68.2%). Focusing further on the comparison of the iterative updating and multi-step test set estimates during the Omicron wave of rounds 17 to 19 (from 5 January 2022 to 31 March 2022), the multi-step estimates suffered severely during these rounds, seemingly due to the variant-induced reduction in population-level faecal shedding and thus, the changes to the prevalence-to-wastewater relationship (Figures 3.1 – 3.2). The poor predictive performance of the multi-step estimates across rounds 17 to 19 is reflected in the accuracy for detecting LTLA-level (52.0%, 95% CI: 48.7%, 55.3%) and regional prevalence movements (48.1%, 95% CI: 28.7%, 68.1%) (Table SI A.8). Such inability to detect prevalence movements is in contrast to the iteratively updating model's estimates where the accuracy at an LTLA level was 75.9% (95% CI: 73.1%, 78.7%) and at a regional level was 81.5%, 95% CI: 61.9%, 93.7%).

Despite the inability of our iteratively trained wastewater model to account for large initial surges in REACT-1 prevalence (in round 17) induced by the Omicron variant, following further model calibration in rounds 17 and 18, in round 19 (the Omicron BA.2 peak), our out-of-sample wastewater-based LTLA-level and regional estimates closely track the movements in LTLA-level and regional REACT-1 prevalence levels, with corresponding prevalence movement detection accuracies of 91.3% (95% CI: 87.5%, 94.2%) and 100.0% (95% CI: 66.4%, 100.0%) respectively (Table SI A.6, Figure 3.2). Indeed, even if there were only a single new model update (or equivalently new calibration round), say round 17 (i.e. omitting REACT-1 round 18), during the Omicron wave (round 19, from 5 January 2022 to 31 March 2022), the wastewater model enabled detection of 87.4% (95% CI: 83.2%, 90.9%) of the within-round LTLA-level REACT prevalence changes.

3.2.3 Complementary Use of WBE for Monitoring SARS-CoV-2 Prevalence

We have thus far demonstrated unstable wastewater-based model performance for prevalence estimation in out-of-sample scenarios without concurrent prevalence surveys, possibly due to the temporal inconsistency of the prevalence-to-wastewater relationship (e.g. due to change in faecal shedding related to different variants of the virus) and/or limitations of our spatiotemporal alignment. We now focus on assessing the utility of WBE in England to complement and enhance prevalence surveys for representative, high-resolution estimation of SARS-CoV-2 prevalence for both economic and logistical efficiency. In short, we no longer omit the survey-based prevalence estimates for a full survey round when testing our model's predictive performance as results are now presented for varying simulated (counterfactual) intensities of community prevalence surveys. To ensure robustness of inferences (about the utility of WBE), we focus primarily on results for the *late* period (rounds 12 to 19, 20 May 2021 to 31 March 2022) due to more complete, representative wastewater surveillance. Our approach here is two-fold as we examine simulated environments i) with reduced geographic survey coverage, and ii) with reduced survey-round sample size of participants.

First, for the simulated scenarios of varying LTLA-level geographic survey coverage, across the varying proportions of LTLAs within each round that were used for model training, to enable equitable comparisons, we specified a random 10%

Table 3.1: **Results of wastewater-based complementary prevalence estimates across rounds 12 to 19 (20 May 2021 to 31 March 2022) using varying geographic coverage of prevalence surveys for training**. Wastewater-based models were trained using 40% to 90% of each REACT-1 round's LTLAs (i.e. 40% to 90% of the SARS-CoV-2 prevalence estimates and of the SARS-CoV-2 wastewater concentrations), and 10% of the LTLAs per round were used for real-time, out-of-sample wastewater-based prevalence estimation. Identical test sets of equal size (n = 247, across eight rounds) were used (within each training-testing environment) to enable comparison of the model-based information offered by various training environments, whilst the procedure was repeated 50 times (i.e. 50 folds) to reduce sensitivity to any individual training-testing splits. The average test set prevalence across the 50 folds was 2.25%. Note that MAE represents the median of Mean Absolute Error values and r represents the median of the fold-specific Pearson's correlation coefficient across the all testing observations *across* the eight rounds, thus measuring modelling performance across sustained time periods (as opposed to performance within individual survey rounds). *Change Detection* indicates the median of the fold-specific directional accuracy, whilst the corresponding 95% CI is attained by Clopper-Pearson method, otherwise known as the Exact Confidence Interval [482].

Training-Testing	MAE	r	Change Detection (95% CI)
40% - 10%	0.69%	0.90	84.2% (79.7%,88.8%)
60% - 10%	0.63%	0.91	85.4% (81.0%,89.8%)
80% - 10%	0.59%	0.93	85.8% (81.5%,90.2%)
90% - 10%	0.56%	0.94	86.6% (82.4%,90.9%)

of withheld LTLAs from each survey round for wastewater-model-based prevalence estimation. Fixed test set sizes across the training sets of different sizes enabled equitable comparisons of model performance and the procedure was replicated 50 times, across different training-testing environments of the same proportions, to reduce sensitivity of inferences to individual training and testing sets.

As expected, relative to survey-based REACT-1 prevalence levels, maintaining a higher geographic coverage of prevalence surveys throughout the ten-month period allowed more accurate out-of-sample wastewater-based prevalence monitoring across the rounds (Table 3.1 and Figure SI A.5). Nevertheless, training environments with fewer LTLA-level observations enabled representative indication of rising or declining REACT-1 prevalence levels for the omitted LTLAs across the sustained period, although the precise estimates tended to depart from survey-based prevalence levels. Similar results were attained for training-testing environments in the *early* period of 24 July 2020 to 3 May 2021 (Table SI A.9), as greater survey geographic coverage (during model training) yielded more accurate wastewater-model-based estimates of test set prevalence, although increasing coverage from 80% to 90% of a REACT-1 round's LTLAs yielded marginally inferior predictive performance (possibly due to incomplete LTLA-level wastewater geographic coverage or the noisiness during epidemic periods with lower SARS-CoV-2 prevalence). The representativeness across rounds of out-of-sample prevalence estimates (over time) for the test set (i.e. omitted) LTLAs, in the presence of a concurrent, reduced-scale prevalence survey, is an apparent consequence of the regular, simultaneous model calibration (which allows for temporal monitoring and thus, updating of the prevalence-to-wastewater relationship). The performance may also be enhanced by the spatial consistency of the prevalence-to-wastewater relationship (Figure SI A.6). For example, in the *late* period with expanded wastewater surveillance, estimated population-level faecal shedding rates generally became increasingly consistent across LTLAs.

Finally, we also evaluated how wastewater surveillance (and our associated modelling) in England could have enabled reductions in the number of participants within survey rounds and thus contributed to greater cost efficiencies in prevalence surveys. Across each setting of varying (simulated) reduced survey-round sample sizes of participants, relative to a (pseudo-independent, ideal) benchmark of spatially smoothed prevalence levels derived from full-sized surveys (see Materials and Methods for further details, including limitations), we observed a consistent trend (Figure 3.3) of wastewater-model-based estimates improving upon the corresponding reduced-survey-based prevalence estimates (in terms of predictive accuracy). Within each round, the average MAE (relative to the aforementioned benchmark derived from full-sized surveys) expectedly declined as the survey round's sample size totals increased for both the reduced-survey-only and combined survey-wastewater environments, although the wastewater model consistently yielded improvements upon the

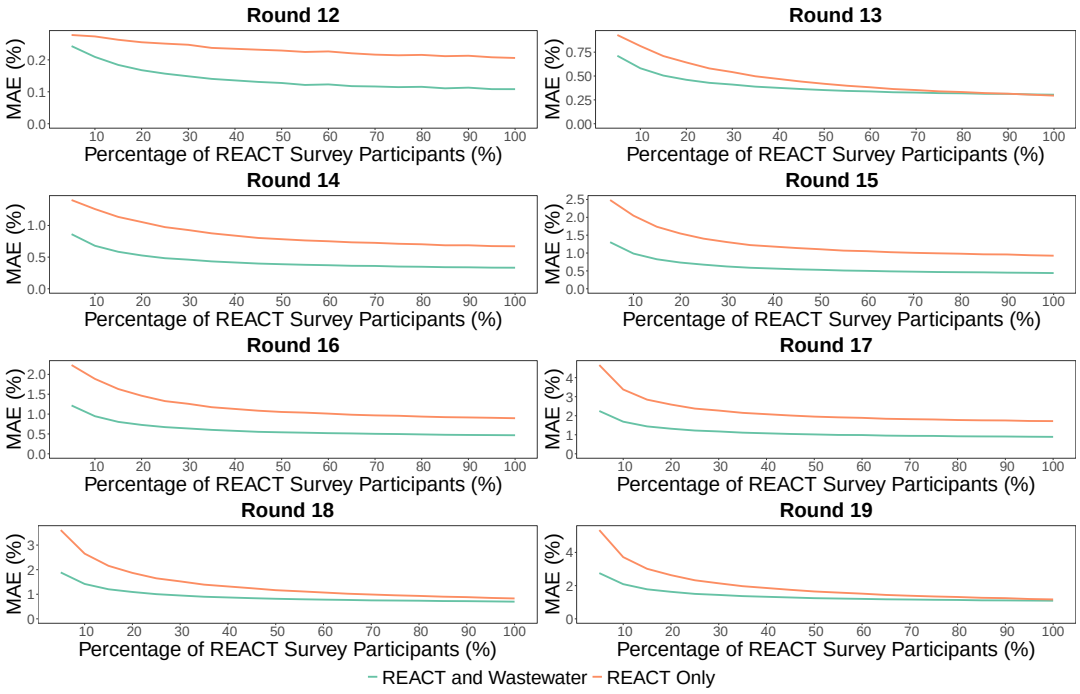


Figure 3.3: **SARS-CoV-2 prevalence estimate accuracy in simulations of reduced survey sample size of participants with and without wastewater modelling.** In individual rounds of 12 to 19 (from 20 May 2021 to 31 March 2022), the following procedure was replicated 100 times for proportions ranging from 5% to 100% in increments of 5%: Environments were simulated to include 5% to 100% of each LTLA's REACT-1 survey rounds' total number of participants, and the number of positives was simulated binomially from each such reduced total number of participants with success probability equal to the REACT-1 weighted prevalence. Then, for each proportion, reduced-sized-survey-based prevalence levels were used to calibrate wastewater-model-based estimates. Both the reduced-survey-based prevalence levels (orange) and combined reduced survey-wastewater estimates (green) were compared, to a spatially smoothed REACT-1 prevalence benchmark derived from full-sized surveys, via the Mean Absolute Error (MAE) within each of the 100 simulation replicates. The mean average of the MAE values (across the replicates) is visualised above. Note the different scales of the y-axes for MAE which are instructive due to the differences in mean prevalence levels (and hence, MAE) across the eight survey rounds.

survey-based estimates for almost every round and survey sample size of participants. Irrespective of whether we attempted to spatially smooth prevalence estimates from reduced survey total numbers of participants (with and without wastewater), wastewater-based modelling improved the accuracy of prevalence estimates obtained from reduced surveys (where accuracy is defined to the spatially smoothed prevalence benchmark derived from full-sized surveys).

3.3 Discussion

Across a time horizon of 21 months of the COVID-19 pandemic in England, our spatiotemporal analyses are the first nationwide detailed evidence synthesis of i) the prevalence-to-wastewater relationship, and relatedly, ii) the utility of WBE for SARS-CoV-2 prevalence estimation. Our investigations uncover how increasing vaccination coverage levels, dynamically changing circulating variants, and evolving population immunity levels are important factors in understanding the prevalence-to-wastewater relationship.

Exploring in further detail the evolution (over time) of correlations between SARS-CoV-2 prevalence in the population and viral particle concentration in wastewater, and the corresponding ratio of prevalence to wastewater concentration, we found strong correspondence between the survey-based prevalence and the wastewater concentrations during the early studied epidemic periods with low prevalence levels and negligible vaccination of the population. Nevertheless, as mentioned above, we subsequently identified two identifiable factors in a temporally evolving prevalence-to-wastewater

relationship; i) levels of vaccination coverage and ii) differences in faecal shedding according to predominant variants. There exists a myriad of potential external influences (or confounders) on the relationship such as changing population-level immunity and human mobility patterns in response to time-varying non-pharmaceutical intervals and the epidemic trajectory[54]. However, reduced population-level faecal shedding following widespread vaccination is consistent with findings from community-based wastewater surveillance of COVID-19 in educational facilities[487]. On the other hand, variant-specific faecal shedding rates (specifically lower shedding rates during the Omicron period) may have led to the observed modifications to the relationship between wastewater concentrations and prevalence. However, the initial breakdowns in the relationship at the beginning of the Omicron wave cannot be disentangled from the greater background vaccination and immunity, compared to previous waves. Indeed, changes in the relationship had already been observed in the preceding months (from September 2021), possibly due to changes to the dominant SARS-CoV-2 variant (as the Delta variant became dominant in the English population)[302]. Nevertheless, the relationship between wastewater concentrations and REACT-1 prevalence estimates is stronger (and more consistent) once again when we restrict measurement focus to the three rounds where Omicron (BA.1 and BA.2) is the dominant SARS-CoV-2 variant, which may be indicative of a returning stabilisation of the prevalence-to-wastewater relationship as the population-level faecal shedding settled at reduced levels. Our findings of implied reduced faecal shedding during the Omicron wave were similarly observed in a community-based analysis of wastewater in the United States[303]. While our findings are at the population (rather than individual) level, they are also consistent with clinical data for the Omicron variant which indicate preferential infection of the upper airway, selective reduction in Omicron infectivity in the intestinal tract, as well as lower viral loads and shorter duration of respiratory shedding, compared to the Delta variant[483,484,485,488].

Leaning now on the results of our modelling analysis, across both periods of our analysis (which are characterised by highly contrasting epidemic dynamics and differing intensities of wastewater surveillance), a trained wastewater-based gradient boosting model with iteratively produced prevalence estimates yielded moderate predictive accuracy for estimating survey-based, high spatial resolution (LTLA-level) prevalence for individual rounds (i.e. up to one month) in settings without any concurrent prevalence survey data. Interpreting these findings alongside the stronger predictive accuracy in the coarser spatial resolution of regional-level, we deduce that reliably estimating SARS-CoV-2 prevalence from a (recently) calibrated wastewater model alone for a single survey round (i.e. for up to one month) is a feasible task.

Whilst wastewater-model-based multi-step prevalence estimates (i.e. produced using wastewater data alone for several individual rounds without any recalibrations of the prevalence-to-wastewater relationship) in the *early* period (24 July 2020 to 3 May 2021) generally provided representative accounts of out-of-sample LTLA-level and regional trends for up to three months (without a simultaneous prevalence survey), there was a notably different picture in the *late* period (20 May 2021 to 31 March 2022). In particular, in the *late* period, characterised by high prevalence levels, increasing vaccination coverage, and changing variants, the multi-step estimates failed to provide an accurate account of the substantial prevalence movements present throughout the Omicron wave, at both LTLA and regional spatial resolutions. Such weak out-of-sample predictive performance (when the wastewater model was not recalibrated for approximately three months) may be a consequence of the increasingly temporally unstable relationship between wastewater and prevalence (apparently influenced by both vaccination and the emergence of the Omicron variant). Our identified lack of temporal consistency is compatible with the findings of Xiao et al.[478], where the wastewater-to-clinical cases ratio also changed substantially during the pandemic. In contrast to multi-step estimates, our iteratively produced wastewater-model-based prevalence estimates for the Omicron BA.2 peak in round 19 (8 March 2022 - 31 March 2022) were strongly representative of the observed sharp increases in prevalence at LTLA and regional level. Indeed, a single additional round of model recalibration during the Omicron wave enabled more representative estimates of survey-based prevalence. These examples highlight the important lesson; accuracy and utility of wastewater-model-based prevalence estimates relies heavily on understanding any changes to the circulating variants and relatedly, the prevalence-to-wastewater relationship. We deduce the importance

of concurrent prevalence surveys for wastewater-based model calibration, alongside the potential use case of WBE which could be undertaken between rounds of a prevalence survey to improve continuity of prevalence estimates, contingent on availability of vaccination and variant tracking data to inform the updating of the prevalence-to-wastewater relationship.

Limitations of our analysis include the sparse geographical coverage of wastewater collection sites earlier in the pandemic, sensitivity of the wastewater surveillance (and the unknown varying sampling strategies) for quantifying RNA concentrations[489], and imprecision of the REACT-1 prevalence survey data at the LTLA level, as well as possible inaccuracies in the weighted estimates employed to correct for variable response rates in different sectors and demographics of the population[308]. Further limitations (discussed in Appendix A.6 and A.8) include the temporal resolutions of our data, the discrete time periods of data collection, the lack of adjustment for variability in estimated wastewater concentrations and prevalence estimates, and the spatial and temporal mismatches which result in averaging of concentrations across treatment plants and across a survey round. Additionally, we acknowledge that our analysis of the utility WBE is strictly confined to the use case within a nation where a community prevalence survey (REACT-1) exists to calibrate the prevalence-to-wastewater relationship.

Focusing now on the complementary role of WBE, by leveraging the estimated spatial consistency of the prevalence-to-wastewater relationship (Figure SI A.6), we show that wastewater-based modelling can improve logistical and economic efficiency of prevalence surveys by providing prevalence estimates over sustained periods for geographies not covered by prevalence surveys. Similarly, relative to spatially smoothed survey-based prevalence, for simulations of reduced survey-round sample size, combined survey-wastewater-based estimates consistently improved upon the accuracy of prevalence levels from reduced surveys alone, even for extremely small-sized surveys. Therefore, we demonstrate that combining WBE with intermittent survey data or fewer samples of survey data may help to maximise the geospatial reach, accuracy, and value of the information, while reducing the cost. Thus, WBE data could be used to provide additional, reliable information to fill in the spatial gaps in estimating SARS-CoV-2 prevalence during intervals with reduced survey data. This could be a cost-effective approach, since the WBE data are relatively cheap to obtain compared to conducting population surveys. Further work including an economic evaluation would be needed to model different scenarios based on timing, extent, and duration of survey data and geospatial coverage of the wastewater data. Whilst this is beyond the scope of the current work, briefly, a cost-benefit analysis for Germany estimated that the national cost for WBE surveillance reagents would be only 0.014% of those required for clinical testing[490], although this did not account for the relatively unknown value (accuracy) of information from WBE compared to prevalence survey data. We recognise that community prevalence studies were not a feature of many nations' response to the COVID-19 pandemic and thus, it is somewhat unclear how WBE could be used to estimate SARS-CoV-2 prevalence in the absence of prevalence surveys in a nation. Whilst beyond the scope of the current research, we tentatively suggest that in this scenario, WBE could be used for well-documented alternative uses (such as early warning detection and variant tracking[303,491]) or concurrent prevalence-wastewater analyses from other nations may be used to calibrate how the wastewater concentrations correspond to community prevalence.

From a public health policy perspective, we have uncovered several important lessons which may be useful for future infectious disease outbreaks. These lessons include that using WBE alone may be insufficient for modelling survey-based SARS-CoV-2 prevalence over several months without any concurrent prevalence survey, potentially hampered by limitations (including wastewater uncertainty, survey-based uncertainty, and/or spatiotemporal mismatches). However, for time horizons up to one month, contingent on recent survey-based calibration of the prevalence-to-wastewater relationship, a wastewater-based model can fill any temporal gaps in surveillance between survey rounds by providing useful, reliable information on prevalence trends in both fine and coarse spatial resolutions. Indeed, reliable wastewater-model-based prevalence estimation, at fine or coarser spatial resolution is often contingent on recent model calibration (via community prevalence surveys), with the regularity of recalibration determined by factors including changing dominant variants and changing vaccination levels. Importantly, wastewater-based modelling can evidently play a complementary role in

improving the cost efficiency of prevalence surveys by also filling gaps in spatial coverage of prevalence surveys or by enhancing accuracy of reduced-size surveys. Thus, there is clear value in the ability of WBE to complement prevalence surveys and achieve greater cost efficiencies. Therefore, we deduce that the appropriateness of WBE in England is dependent on the precise use case (WBE alone or complementary to prevalence surveys), the spatial resolution (coarse or fine), and the concurrent epidemic dynamics (such as vaccination and variants). In the event of future outbreaks of infectious diseases, a combination of surveillance by survey data and WBE is likely to be a cost-effective approach to obtaining situational awareness for policy makers. This will likely require concomitant monitoring of vaccination coverage and variants (accessible via prevalence surveys, WBE, and genotyping), alongside clinical indicators of epidemic dynamics. Together, these will help to determine the optimal timing of wastewater model recalibration and the scale of survey and wastewater monitoring required to ensure representative prevalence estimation.

3.4 Materials and Methods

3.4.1 Study Design

The objectives of the study were twofold; i) to establish how SARS-CoV-2 wastewater concentration estimates related to corresponding SARS-CoV-2 community prevalence estimates and ii) how wastewater concentrations could be effectively used, possibly alongside community prevalence surveys, for monitoring SARS-CoV-2 prevalence in varying spatial resolution across England.

Wastewater concentration data were sourced from the Environmental Monitoring for Health Protection (EMHP) programme, which was led by the United Kingdom Health Security Agency (UKHSA) and tested untreated sewage across England for fragments of SARS-CoV-2. The programme monitored wastewater viral concentrations of SARS-CoV-2 RNA, variants of concern, and variants under investigation. For estimation of wastewater concentrations, samples were collected three to four times weekly at sewage treatment works (STWs) and sewer network sites[302]. The monitoring programme commenced in July 2020 at 45 STWs and a large-scale expansion occurred nationally in June 2021. The wastewater samples were obtained in one of two ways: either via i) grab samples (a single sample taken using a small container at one point in the day), or ii) composite samples (an autosampler samples at regular intervals throughout the day and mixes the samples together in a container). The programme did not disclose details of the method used (grab or autosampler) for individual samples. See Appendix A.8 for discussion of further sources of wastewater uncertainty and potential inaccuracies[302,489].

Viral concentrations of SARS-CoV-2 RNA were obtained from wastewater samples by quantifying the number of SARS-CoV-2 N1 gene copies per litre via the Reverse Transcriptase Polymerase Chain Reaction (RT-qPCR) process, a quantification method which combines two main steps of reverse transcription (RT) and quantitative PCR (qPCR)[294]. Normalisation of wastewater flows is necessary to account for precipitation dilution of concentrations of SARS-CoV-2 RNA. The indirect normalisation conducted under the EMHP programme is described elsewhere[489], but briefly, the technique assumed that wastewater flows were not directly observable, flow variability was estimated and outliers were identified. The data were then presented as the flow-normalised viral concentrations of SARS-CoV-2 RNA, defined in units of numbers of gene copies per litre.

Community prevalence data were sourced from the REACT-1 study which obtained estimates of weighted prevalence of SARS-CoV-2 infection in England from 1 May 2020 to 31 March 2022. Across 19 distinct rounds of cross-sectional surveys[309], random samples of the English population (over 5 years of age) were taken[307,308]. Rounds of REACT-1 were approximately monthly with round durations between 15 and 31 days. The study involved participants using a self-administered throat-and-nose swab kit and completing a questionnaire. The swabs were sent to a laboratory (either by courier or priority post) where the samples were tested for SARS-CoV-2 by RT-PCR. The laboratory extracted nucleic

acid for SARS-CoV-2 using the ViroBOAR 1·0 RT-qPCR kit (EuroFins Genomics, GmbH, Ebersberg, Germany) and the Roche Lightcycler 480 II (Roche Diagnostics, Almere, The Netherlands) to detect in parallel 2 gene targets: the N gene and E gene. Valid laboratory test results were defined as positive for an individual if both the N gene and E gene targets were detected or if the N gene was detected with a cycle threshold (Ct) value less than 37. Date of swabbing was used for prevalence estimates where available, and otherwise the study used date of pickup by courier or date of scan in post office[307,308].

Throughout the 19 rounds, there were approximately 2.5 million respondents who had a valid laboratory RT-PCR test for SARS-CoV-2. The use of random sampling of individuals in the study was advantageous as it avoided many of the inherent biases that exist with only testing of symptomatic individuals, test-seeking behaviours, and the availability of tests. Nevertheless, the REACT-1 study also used a within-round random iterative method (RIM)-weighting to correct prevalence estimates to be representative of the population (and thus, account for possible non-response bias)[492]. The weighting adjusted for gender, age, LTLA counts, deprivation status, and ethnic group[307]. We provide more extensive details of the REACT-1 study design, sampling strategy, and statistical analyses in Appendix A.7.

3.4.2 Geospatial Mapping of Wastewater Concentrations and Relating to REACT-1

In our analysis, our temporal resolution is individual REACT-1 survey rounds. For each survey round, we took the median value of an STW's flow-normalised viral concentrations (henceforth termed the *concentrations*), which are then log-transformed. In terms of spatial resolution, the catchment areas of STW locations for the reported concentrations do not align with the spatial resolution of the REACT-1 study; the 315 LTLAs of England. We therefore used a geospatial approach to align the spatial resolutions i.e. to map the wastewater concentrations from the STW level to LTLA level (of the prevalence survey). We employed the lookup table provided by[493], which gives a mapping of wastewater catchment areas to Lower layer Super Output Areas (LSOAs); small regional geographies that combine to form the larger geography of an LTLA. For 21 STWs of the EMHP programme which are not included in the lookup table, to estimate their catchment areas, we employed an approximation method proposed by[480].

Equipped with the catchment-to-LSOA mapping area data and Office for National Statistics (ONS) 2019 mid-year population estimates, we then derived geospatial population estimates (GPEs) for the intersection area between each catchment area of STW i and LTLA k. Our geospatial approach is adapted from[493], and we assumed that each LSOA population is uniformly distributed across the area serviced by STWs:

$$\widehat{P_k} = \sum_{j=1}^{J} \widehat{p_j},$$

where $\widehat{P_k}$ is the GPE for LTLA k and $\widehat{p_j}$ is our estimate of the serviced population of LSOA j based on our population estimates for the intersection area between the catchment area of STW i and LSOA j (see Appendix A.6 for further details).

For each LTLA k, the weight w_{ik} assigned to STW i for concentrations was the ratio of the GPE for the intersection area between the catchment area of STW i and area of LTLA k divided by the GPE for LTLA k:

$$w_{ik} = \frac{\sum_{j=1}^{J} \widehat{g_{ij}}}{\widehat{P_k}},$$

where $\widehat{g_{ij}}$ are the GPEs for the intersection area between the catchment area of STW i and area of LSOA j. If any particular STW s failed to report concentrations within a round of the REACT-1 study, the weights of LTLA k for that round were re-weighted such that $\sum_{i \neq s} w_{ik} = 1$ and $w_{sk} = 0$.

Finally, for a given round, the estimated concentration of an LTLA was the weighted average of the median STW-level concentrations within that LTLA. The weights employed here were the STW-LTLA population estimates. We also

considered alternative approaches such as taking the wastewater measurement of the (single) dominant STW in an LTLA, the mean STW-level concentration across the round (as opposed to the median), and various lead/lag times where we expanded/reduced the window for wastewater measurements to account for potential early or prolonged wastewater-based detection of SARS-CoV-2 (Appendix A.9). Each alternative approach did not led to an enhanced relationship between the estimated wastewater concentrations and REACT-1 prevalence estimates.

In the event of future outbreaks of infectious diseases, the developed approach could be applicable to population-level wastewater analyses which require spatial alignment of wastewater catchments to geographies used by public health authorities. Nevertheless, there exists limitations with the predictive approach of using GPEs, such as transient populations, uncertainty surrounding population estimates of small geographic areas, and potentially unrepresentative wastewater concentration estimates for some LTLAs, each of which are discussed in Appendix A.6.

3.4.3 Relationship Between Wastewater Concentrations and REACT-1 Prevalence

Our spatiotemporal analyses are divided into two distinct periods; the *early* period of rounds 3 to 11 (24 July 2020 to 3 May 2021 - the first time periods where the EMHP programme intersects with the REACT-1 study) and the *late* period of rounds 12 to 19 (20 May 2021 to 31 March 2022). The two periods are primarily analysed independently due to differences in wastewater surveillance geographic coverage and the EMHP programme's transition from usage of two laboratories to a single laboratory.

To allow for potential nonlinearities, Spearman's correlation was used to assess the relationship between (log) concentration and SARS-CoV-2 prevalence. Note that log concentration, as opposed to concentration, was used in these analyses as this variable served as our predictor within our best-performing wastewater-based gradient boosting models (see below). In particular, we studied correlations between all LTLA-level (and regional population-weighted averages of) SARS-CoV-2 wastewater concentrations estimates and REACT-1 SARS-CoV-2 prevalence estimates across i) the *early* and *late* periods separately, and ii) across moving five-round windows centred on each survey round from REACT-1 rounds 5 (starting 18 September 2020) to 17 (ending 20 January 2022). Then, for comprehensiveness, we also assessed the correlations within the time series for each individual LTLA (and each region) to understand any variability in the relationships for different geographies nationally.

3.4.4 Modelling REACT-1 SARS-CoV-2 Prevalence

Our primary model for estimating LTLA-level SARS-CoV-2 prevalence is an extreme gradient boosting model which is implemented via the `xgboost` package in R version 4.2.1[494,495]. Briefly, the extreme gradient boosting (known as XGBoost) machine learning algorithm is comprised of an ensemble of weak learners consecutively fit to data in a greedy manner. The algorithm is a parallelised, optimised version of the general gradient boosting algorithm[496]. XGBoost was chosen due to its predictive capabilities and highly flexible nature (ability to handle complex nonlinearities and time-varying relationships such as the prevalence-to-wastewater relationship over the course of the COVID-19 pandemic), whilst also enabling regularisation (via model hyperparameters such as the learning rate) and cross-validation. Indeed, XGBoost often produces state-of-the-art predictive performance across a variety of settings, including for our research objective of estimating SARS-CoV-2 prevalence from wastewater concentrations[479]. We developed a Bayesian hierarchical model to ensure robustness of conclusions from our modelling analysis (described in Appendix A.10). Vaccination data from REACT-1 were used as covariates in the *late* period, and a full description of covariates used in our wastewater-based models of SARS-CoV-2 prevalence is provided in Table SI A.2. 95% prediction intervals (PIs) were obtained using 5,000 non-parametric bootstrap samples/replicates of the original training sets.

3.4.5 Types of Wastewater-Model-Based Estimates

Our appraisal of model performance focused on wastewater-model-based estimates of SARS-CoV-2 prevalence. Note that we only trained our model at an LTLA spatial resolution (as model training at a regional level produced inferior accuracy in regional estimates to those obtained by training at an LTLA level). Regional estimates were achieved by weighting the LTLA-level estimates by the corresponding LTLA populations within the region.

There were three broad types of scenarios for wastewater-model-based prevalence estimates: i) *Iterative training and testing* involved calibrating our wastewater-based model using a minimum of four REACT-1 rounds and subsequently using wastewater data in our model to estimate prevalence for a full REACT-1 round (for which we omitted prevalence data) one-at-a-time. In further detail, we made use of wastewater and prevalence data up to one survey round before each testing round, and then our wastewater-based model used the testing round wastewater data to produce prevalence estimates in each testing round. The idea was to evaluate inferences gained from subsequently relying on WBE alone (for up to one month) or temporarily between rounds of a prevalence survey. ii) *Multi-step testing* involved training our wastewater model for five to six survey rounds but using wastewater data to estimate prevalence for three individual survey rounds (for which prevalence data was omitted for testing) without recalibrating the model. In other words, the model learned the relationship between wastewater and prevalence during training rounds, and then used wastewater data alone (within the model) during model testing for estimation of prevalence within the three individual survey rounds. The intuition here was to appraise the exclusive application of WBE over an approximate three-month period without any concurrent REACT-1 survey data to recalibrate the prevalence-to-wastewater relationship. iii) *Complementary wastewater-based estimation* differed to the cases above as it did not omit all of the testing rounds' prevalence data and hence, involved partially (counterfactually) complete, survey rounds. In doing so, we assessed the effectiveness of WBE for complementing incomplete prevalence surveys which could arise due to funding and/or logistical limitations which prevent a large-scale, national survey programme. Here, we a) first considered reduced geographic coverage of prevalence surveys as we excluded varying numbers of random LTLAs from single survey rounds, and the corresponding prevalence levels were estimated using our wastewater-based model and compared to corresponding REACT-1 prevalence estimates. b) Second, we investigated how WBE could contribute to prevalence estimation in the event of reduced survey sample size totals (i.e. fewer survey participants) within rounds. Using reduced survey-round totals of participants, we simulated the number of positive individuals in an LTLA (within a round) from a binomial distribution with success probability equal to the REACT-1 weighted prevalence. We replicated the simulation 100 times to reduce sensitivity to sources of randomness. A key challenge in the simulation analysis was to decide an appropriate benchmark against which we could compare accuracy of prevalence estimates, in survey rounds of varying sizes, with and without a wastewater model. The lack of an external benchmark to validate estimates is a key limitation of our simulated reduced sample size analysis. In the absence of knowing the true prevalence in communities (the ideal benchmark), we used a pseudo-independent, spatially smoothed version of reported REACT-1 prevalence levels (i.e. spatially smoothing the full-sized survey-based prevalence estimates). In doing so, we attempted to reduce the likelihood of bias and misleading conclusions potentially caused by benchmarking reduced-survey estimates against either complete survey-based prevalence estimates or completely wastewater-model-based estimates, whilst further considering the potential presence of spatial autocorrelation in SARS-CoV-2 prevalence. Briefly, the degree of spatial smoothing of LTLA-level prevalence observations within each round was determined by fitting of simple Kriging models to each of the 309 observations and the distances between LTLA centroids, thus accounting for the degree of spatial autocorrelation in prevalence levels within each survey round. Sensitivity of inferences to the choice of spatial smoothing method was appraised by also smoothing according to a smoothing kernel.

Due to our study's objective of assessing the utility of WBE for complementing or replacing survey-based prevalence estimates, the predictive accuracy of wastewater-model-based estimates (i.e. performance of our model on new, unseen data) were appraised relative to REACT-1 SARS-CoV-2 prevalence estimates (unsmoothed unless stated otherwise), using metrics

such as MAE, Pearson's correlation with the response (r), and accuracy in detecting directional prevalence movements. As MAE is generally dependent on the prevalence levels of a specific round, we often report the mean prevalence level to provide context for each attained MAE. Note that we also assessed LTLA-specific predictive accuracy by focusing on the estimates for each LTLA/region individually (i.e. stratifying the time series) and then reporting summary statistics across the LTLA/region-specific performance metrics. The intuition for assessing LTLA/region-specific predictive accuracy is that we may be interested in tracking the prevalence trends of individual geographies over time.

3.4.6 Data Availability

Access to REACT-1 individual-level data is restricted to protect participants' anonymity. LTLA-level REACT-1 data can be obtained under data request. The wastewater concentration surveillance data from the EMHP are publicly available at `https://www.gov.uk`. The wastewater catchment areas for the United Kingdom are available at `https://github.com/tillahoffmann/wastewater-catchment-areas`.

Chapter 4

Unifying techniques for climate-informed dengue modelling and forecasting

The following chapter presents two related manuscripts[45,46] which introduce new methods for climate-informed modelling of past, current, and future (short-term) dengue epidemic dynamics. The methods are applied, in both manuscripts, to dengue epidemics in northern Peru – a region that has been at the epicentre of recent dengue emergencies.

4.1 Paper I: Climate-based modelling and forecasting of dengue in three endemic departments of Peru

Cathal Mills and Christl A. Donnelly. Climate-based modelling and forecasting of dengue in three endemic departments of Peru. *PLOS Neglected Tropical Diseases*, 18(12):e0012596, December 2024. ISSN 1935-2735. doi: 10.1371/journal.pntd.0012596. URL https://dx.plos.org/10.1371/journal.pntd.0012596

This first manuscript, published in *PLOS Neglected Diseases*[45], introduces methods for probabilistically quantifying the roles of multiple climatic influences on dengue incidence, which then informs a forecasting model for future dengue incidence.

Abstract

Amid profound climate change, incidence of dengue continues to rise and expand in distribution across the world. Here, we analysed dengue in three coastal departments of Peru which have recently experienced public health emergencies during the worst dengue crises in Latin American history. We developed a climate-based spatiotemporal modelling framework to model monthly incidence of new dengue cases in Piura, Tumbes, and Lambayeque over 140 months from 2010 to 2021. The framework enabled accurate description of in-sample and out-of-sample dengue incidence trends across the departments, as well as the characterisation of the timing, structure, and intensity of climatic relationships with human dengue incidence. In terms of dengue incidence rate (DIR) risk factors, we inferred nonlinear and delayed effects of greater monthly mean maximum temperatures, extreme precipitation, sustained drought conditions, and extremes of a Peruvian-specific indicator of the El Niño Southern Oscillation. Building on our model-based understanding of climatic influences, we performed climate-model-based forecasting of dengue incidence across 2018 to 2021 with a forecast horizon of one month. Our framework enabled representative, reliable forecasts of future dengue outbreaks, including correct classification of 100% of

all future outbreaks with DIR ≥ 50 (or 150) per 100,000, whilst retaining relatively low probability of 0.12 (0.05) for false alarms. Therefore, our model framework and analysis may be used by public health authorities to i) understand climatic drivers of dengue incidence, and ii) alongside our forecasts, to mitigate impacts of dengue outbreaks and potential public health emergencies by informing early warning systems and deployment of vector control resources

Author Contributions

- Conceptualisation: **C.M**
- Methodology: **C.M**; C.A.D.
- Data curation and visualisation: **C.M**
- Writing original draft: **C.M**
- Writing final draft: **C.M**; C.A.D

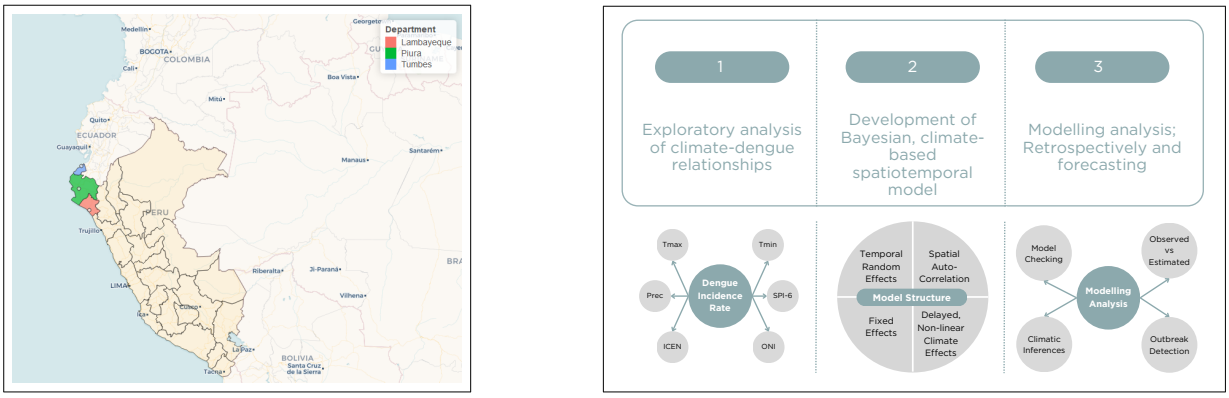
4.1.1 Introduction

In recent years, various analyses globally have quantified complex interactions between climate variability and the dynamics of vector-borne infectious diseases, such as dengue, malaria, Zika, and chikungunya[78,82,88]. Dengue virus (DENV) is an example of a flavivirus that is transmitted to humans by infected *Aedes* mosquitoes (primarily *Ae. aegypti*). Endemic in over 100 countries, dengue continues to pose a substantial public health concern worldwide, particularly in tropical and subtropical regions where the *Ae. aegypti* mosquito thrives. Incidence of dengue cases has risen significantly in recent decades, alongside an expanding geographical distribution that is attributed to factors that regulate transmission dynamics such as urbanisation, globalisation, increased travel, inadequate vector control measures, and climate change[77,352].

Here, we quantify the important role played by the climate in the timing and intensity of reported dengue transmission across endemic regions of Peru. In many other nations, complex, nonlinear associations are well-established between incidence of dengue cases and climatic variables such as temperature and precipitation[78,214]. Climatic variables directly influence the abundance, distribution, and activity of *Ae. aegypti* mosquitoes, as well as the replication and survival of DENV within the mosquito vector. In particular, warmer temperatures accelerate mosquito development, shorten their incubation period, and increase their biting rates. Meanwhile, rainfall generally provides breeding sites for *Ae. aegypti* mosquitoes, promoting their population growth[80,81,82,392,497,498].

Extreme climatic events, such as drought and heavy rainfall, have complex, spatially dependent relationships with dengue transmission. Whilst drought can reduce dengue transmission by limiting mosquito breeding sites and mosquito populations, water scarcity can conversely prompt individuals to store water in containers that serve as mosquito breeding sites, thus producing conditions that favour the survival and proliferation of mosquitoes[78,84,499,500]. Similarly, during periods of intense rainfall, water typically accumulates in containers such as flower pots and water storage containers, thus providing ample breeding grounds for mosquitoes[501,502,503]. In terms of human factors, heavy rainfall can result in the overflow of sewage systems, increasing the risk of contamination and facilitating the spread of dengue.

In Peru, dengue has long been a significant public health concern, characterised by periodic outbreaks and year-round transmission. In recent years, there have been major dengue outbreaks where annual reported incidence has reached record levels within the nation. The studied departments of Piura, Tumbes, and Lambayeque are located in the northern coastal region of Peru, and were chosen in this study due to their presence at the epicentre of recent severe dengue outbreaks[504]. We were interested in quantifying climatic influences on dengue epidemic dynamics in the region as the three departments experience favourable environmental conditions for *Ae. aegypti* mosquito breeding due to their warm and humid climate,



A B

Figure 4.1: **Study setting and schematic of methodology.** **A**: Geography of Peru (shaded amber), neighbouring countries and the Pacific Ocean. The three studied departments of Lambayeque, Piura, and Tumbes are highlighted in red, green, and blue respectively, and their capital cities are represented by filled circles. The map was created using the mapview and rgeoboundaries packages in R version 4.2.1 [495,507,508]. The base layer map is available at `https://www.geoboundaries.org/api/current/gbOpen/PER/`. **B**: Overview of the approach taken for analysis of dengue incidence in the three departments across 2010 to 2021.

whilst coastal proximity increases exposure to extreme El-Niño-induced climatic events. Piura, Tumbes, and Lambayeque have also faced challenges in dengue control due to population growth, sanitation infrastructure, and resources for vector control programmes.

Understanding the climatic influences on dengue transmission in the three departments is essential for development of effective strategies to prevent and control disease transmission. By characterising the relationships between climatic variables and dengue incidence, public health authorities can attain insights for targeted interventions, early warning systems, and resource allocation, thus potentially mitigating the ongoing, persistent public health impact of dengue.

4.1.2 Methods

The following section describes the data, study area, and modelling techniques used to model reported dengue cases in the Peruvian departments of Lambayeque, Piura, and Tumbes from 2010 to 2021.

Data

We obtained department-level surveillance data on reported monthly dengue cases (which includes both probable and confirmed cases across dengue serotypes) from the National Centre for Epidemiology, Disease Prevention and Control (Peru CDC) in Peru's Ministry of Health [505]. The Peru CDC define a probable dengue case for an individual who i) has febrile illness for at most seven days, ii) has two or more of a list of specified symptoms, iii) and lives in or has recently visited areas with either known dengue transmission or known *Ae. aegypti* populations. Confirmed dengue cases meet the same exposure and symptoms criteria, alongside a positive dengue test [504]. We sourced mid-year population estimates (as of 30 June) for each of the three departments from the INEI (National Institute of Statistics and Information of Peru), and for each month, we linearly interpolated the yearly population values to avoid improbable population fluctuations [506].

In terms of meteorological data, we sourced monthly precipitation and temperature data from the WorldClim 2.1 dataset, which provides open-source climate data globally in fine spatial resolution of 1km [281,509]. We derived monthly estimates of average daily maximum temperature (°C), average daily minimum temperature (°C), and total precipitation (mm) for each studied department by averaging the variables across the grid cells corresponding to each department.

We used the Standardised Precipitation Index at a six-month time scale (SPI-6), a widely used index in hydrology and climatology, to assess and monitor drought conditions based on precipitation data [510,511]. SPI-6 measures how anomalous

the accumulated precipitation over six months is compared to historical records, and accounts for both the magnitude and duration of precipitation deficits or surpluses. As SPI-6 is calculated over a medium-term accumulation period, it is an indicator of reduced stream flow and reservoir storage. An advantage of SPI-6 is that is a relative index as it represents the deviation from the average conditions specific to a location. Positive SPI-6 values indicate wetter-than-average conditions, whilst negative values indicate drier-than-average conditions. The SPI-6 data were sourced from a dataset collated by the European Drought Observatory[512].

Finally, for characterising the impacts of the El Niño Southern Oscillation (ENSO) on human cases of dengue, we considered two climate indices of the Pacific Ocean; the El Niño Coastal Index (ICEN) and the Oceanic Niño Index (ONI). The data on the indices were sourced respectively from the Geophysical Institute of Peru (IGP) and the National Oceanic and Atmospheric Administration in the United States. First, the ICEN, maintained by the IGP, is a climate monitoring index which is derived from the three-month running average of the sea surface temperature (SST) anomaly in the Niño 1+2 region and the climatology of the period between 1981 and 2010[513,514]. As the ICEN is a bespoke index that is specific to Peru, it is generally more representative of ENSO conditions for Peru than large-geographic-scale monitoring metrics (such as the ONI). Here, due to our geographic focus on the Peruvian coast, we concentrated on the E-Index component of the ICEN which summarises the variability associated with El Niño (warming events) and La Niña (cooling events) via anomalous surface warming in the eastern Pacific Ocean. Nevertheless, we also extracted values of the ONI, which provides a three-month rolling average of the SST anomaly in the Niño 3.4 region[515]. ONI values greater than 0.5 are indicative of El Niño events, whilst ICEN E-Index values greater than 1.7 are classified as strong El Niño events.

Study area

The three departments of Piura, Tumbes, and Lambayeque are located in the northern coastal region of Peru (Figure 4.1). As of June 2021 (the latest estimates for the studied period), the populations of Piura, Tumbes, and Lambayeque were estimated to be approximately 1,326,000, 2,077,000, and 256,000 respectively (Table SI B.2), whilst the corresponding population percentages living in urban areas were 79.3%, 93.7%, and 81.1%. The climates of each department are characterised by generally dry and hot conditions, as well as distinct dry and wet seasons. The neighbouring departments experience broadly similar climatic conditions (Figure SI B.7), and also share similarities in seasonal trends for dengue incidence rate (DIR) per 100,000 population (Figure 4.2 and Figure SI B.8)[516,517]. Whilst the departments have historically differed in the average DIR observed, all three departments have experienced recent endemicity of dengue and have been afflicted by severe dengue outbreaks in 2015 and 2017 and more recently in 2023 and 2024 (which are not part of the current studied period).

Modelling framework

We developed a bespoke, climate-based Bayesian spatiotemporal model which was heavily shaped by region-specific knowledge of the Peruvian coastal climate and its interplay with dengue incidence. Our underlying model structure was informed by previous climate-based spatiotemporal analyses of dengue incidence in other nations[78,214]. In particular, our Bayesian hierarchical model of monthly dengue incidence allowed for temporal random effects, spatiotemporal effects (both structured and unstructured), and delayed, potentially nonlinear effects of climatic variables. We employed domain knowledge and insights from our extensive exploratory analysis of the climatic influences which allowed us to identify factors specific to Piura, Tumbes, and Lambayeque, as well as initial insights into the strength and nature of the relationships between the climate and reported cases. In our exploratory analysis, we described similarities across departments in the structure and timing of relationships between climatic variables and DIRs (Figures SI B.1 and SI B.7). Alongside the random effects and climatic variables, our model incorporated a common seasonality indicator for the Peruvian summer months and an innovative momentum indicator, the lagged value of the Relative Strength Index (RSI), which was inspired

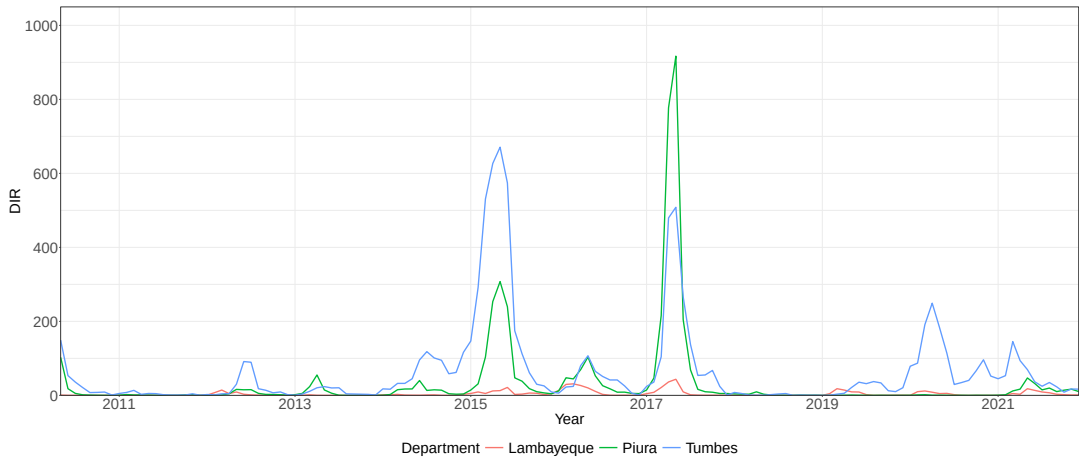


Figure 4.2: **Dengue incidence rate (DIR) trends over time across the departments.** The monthly DIR per 100,000 is displayed for each of the three studied departments across 2010 to 2011 inclusive. Substantial peaks in monthly DIR occurred in 2015 and 2017.

by analysis techniques of financial markets. The epidemiological logic behind the RSI is that it approximates the epidemic trajectory and thus aims to assist in tracking and forecasting temporal trends in incidence of new dengue[518]. Whilst a brief description of the model framework is provided here, further descriptions are provided in Appendix B.3.

To allow for potential overdispersion, we assumed a negative binomial likelihood for the monthly case counts of dengue y_{it} for each department ($i = 1, 2, 3$) across the studied months ($t = 1, \ldots 140$):

$$\begin{aligned} y_{it} &\sim NegBin(\mu_{it}, \kappa) \\ \log(\mu_{it}) &= \log(P_{it}) + \log(\eta_{it}) \\ \log(\eta_{it}) &= \beta_0 + \boldsymbol{x_i}\boldsymbol{\beta} + \gamma_{i,m(t)} + \delta_{i,a(t)} + u_i + v_i + \sum_{l=1}^{L} bs(c_{lit}, 2, 4) \end{aligned} \tag{4.1.1}$$

where the negative binomial distribution had mean parameter μ_{it} and overdispersion parameter κ. We included a population offset $\log(P_{it})$, where log denoted the natural logarithm, such that η_{it} represented the DIR per 100,000 population. We then modelled the DIR with an intercept β_0, alongside fixed and random effects. $\boldsymbol{x_i}$ (with corresponding parameter vector $\boldsymbol{\beta}$) represented a matrix of fixed effects; the binary seasonality variable (1 for summer months of December to April, and 0 otherwise) common across departments and the department-specific RSI variable lagged by one month. To account for cyclical patterns of monthly temporal dependence, we used monthly random effects ($\gamma_{i,m(t)}$) specific to each department, which were specified by cyclic RW(1) (random walk of order one) prior distributions. Furthermore, to account for unobservable year-to-year heterogeneities, such as changes to reporting rates, vector control programmes, or population-level immunity, we included exchangeable (Gaussian-distributed independent and identically distributed) department-specific yearly random effects $\delta_{i,a(t)}$. Sensitivity analyses revealed that pooled monthly or annual random effects were too primitive to adequately capture the departments' monthly and annual variations in DIRs respectively. u_i and v_i represented the structured and unstructured spatial random effects respectively of our a reparameterisation of the traditional BYM (Besag, York and Mollié) model, called the BYM2 model, which implements Penalised Complexity prior distributions to enable scaling and provide interpretable parameters[519,520]. Finally, to allow for varying lagged and potentially nonlinear effects of maximum temperature, precipitation, SPI-6, and ICEN, we used distributed lag nonlinear models (DLNMs) for our four ($L = 4$) climatic covariates c_{lit} via quadratic B-splines $bs(c_{lit}, 2, 4)$ with two degrees of freedom in the exposure-response dimension, a maximum lag of four months, and an internal knot at two months in the lag-response dimension (further details in Appendix B.3.5)[394].

The model was implemented using Integrated Nested Laplace Approximation (INLA) in R version 4.2.1[459,495]. In

addition to assessments of model structure, we specified our final model by analysing performance across diagnostics such as Deviance Information Criterion (DIC), cross-validated (CV) log score, Mean Absolute Error (MAE), posterior predictive checking, and Probability Integral Transform (PIT) values[521].

Modelling analysis

We divided our modelling analysis into three environments; i) model fitting to the entire dataset of 140 months of dengue notifications across the three departments, ii) retrospective modelling with leave-one-out cross-validation (LOOCV), and iii) model-based forecasting with a forecast horizon (or lead time) of one month.

First, our model fit to the entire dataset was the basis for drawing conclusions about the climatic relationships with new dengue cases. We substantiated conclusions by extensive sensitivity analyses where we varied the model structure (B.4.2). Second, in our LOOCV approach (more precisely a leave-one-time-point-out approach) we excluded each time point's (three) monthly observations one at a time, refitting the model to a reduced dataset of 417 observations (139 time points) and generating posterior predictive distributions for each of the three excluded observations. We assessed model performance on the out-of-sample observations via outputs such as the proportion of DIR observations within the 95% credible intervals of the posterior predictive distributions. In doing so, we (in part) externally validated our model's predictive performance in synthetic out-of-sample environments to reduce the likelihood of overfitting. In the LOOCV environment, we further generated posterior probabilities that a monthly DIR exceeded specific thresholds (such as 50 per 100,000) by computing the proportion of each observation's posterior predictive samples which exceeded the corresponding threshold. We assessed our model's ability to accurately detect outbreaks of various sizes in the out-of-sample settings via visualisations of the posterior probabilities (compared against the observed DIR values) and accuracy metrics for classification of outbreak events (predicted if posterior probability exceeded an outbreak classification cut-off probability, defined below). Note that our public-health-oriented exceedance thresholds (for outbreak months) were pre-specified by us to indicate different magnitudes of dengue outbreaks.

Probabilistic forecasting involved the exclusion of an individual month's data and all future data (both climatic and epidemiological) from the model fitting procedure, and then using only data with a lag of one month or more to forecast the specified individual month's DIRs. In the forecasting setting, we focused on 2018 to 2021, a period of four years (one month at a time) to ensure that we had sufficient (surveillance and climate) data for training a well-informed model for reliable forecasting. We performed similar assessments to the LOOCV environment, including posterior predictive 95% credible interval checks and outbreak posterior probabilities. For classification of outbreaks, we did not employ an arbitrary cut-off posterior probability as we used our retrospective modelling analysis to identify an appropriate cut-off probability. In particular, using only historical data from 2010 to 2017, our optimisation procedure identified a cut-off probability (for each threshold of 50 per 100,000 or 150 per 100,000) that historically maximised the percentage of true outbreaks detected, whilst simultaneously minimising the probability of false alarm. Whilst the calibrated outbreak detection performance for the LOOCV historical analysis is an upper bound on historical model performance, for the important use-case setting of real-time forecasting, the cut-off was calibrated using past data only. The procedure was equivalent to maximising the historical AUC, the area under the Receiver Operating Characteristic (ROC) curve. Thus, we ensured that our outbreak cut-off probability was informed by i) priorities of public health authorities (i.e. reliable, robust forecasting of outbreaks) and ii) historical results. Overall, the objective of the forecasting analysis was to determine whether climate-based forecasting could be useful for real-time prediction of future trends in incidence of new cases and reliable detection outbreaks of varying magnitudes one month ahead of time.

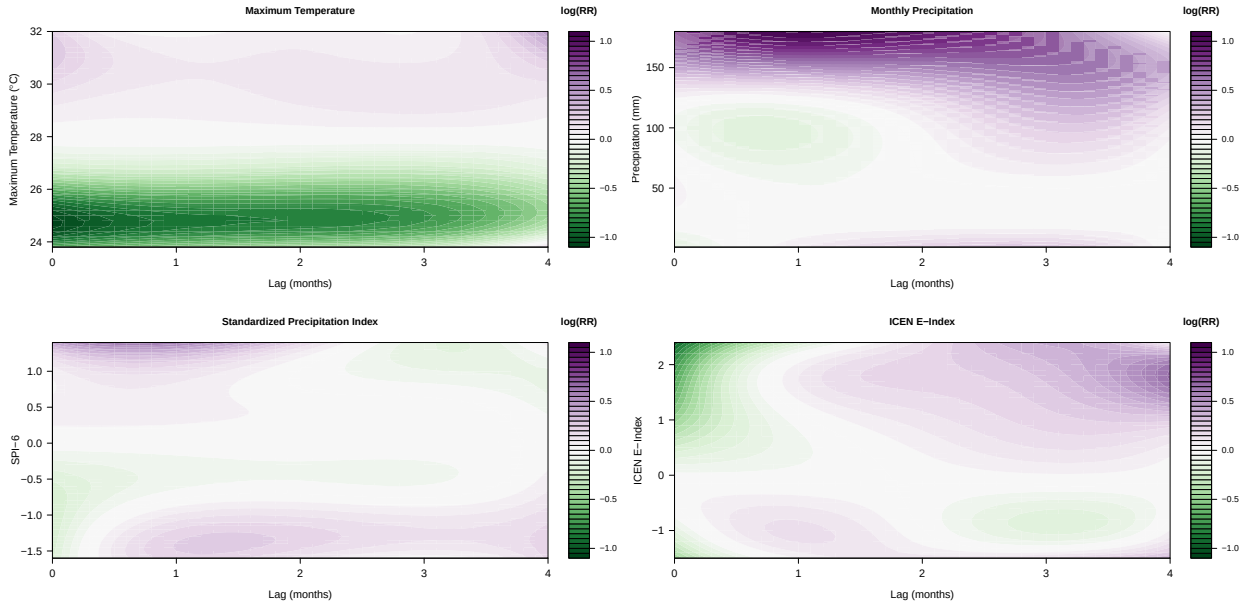


Figure 4.3: **Model-based exposure-lag-response relationships between climatic variables and dengue incidence rates.** Plots of relative risk (RR), on a natural logarithm scale, for the included DLNMs (Distributed Lag Nonlinear Models) in our climate-based Bayesian hierarchical model for incidence (of new dengue cases) fitted to the entire period of 140 months, where RR is defined relative to the risk induced by the mean observed value of each climate variable. Log RR values greater than 0 (pink to purple) correspond to heightened RR of incidence of new dengue cases, whilst values less than 0 (green) correspond to reduced RR. The four climatic variables were included in the model framework via DLNM specifications alongside temporal random effects, spatiotemporal random effects, and fixed effects (of momentum and seasonality).

4.1.3 Results

Characterising climatic influences on incidence of new dengue cases

The exploratory cross-correlation plots (Figure SI B.1) capture the structure and intensity of the linear relationships between lagged values of climate variables and the current DIR. Here, we identified strong similarity in the nature of each department's relationships between climatic factors and DIRs. From a forecasting perspective, the climatic relationships with DIRs were generally strongest (magnitude of correlation) at lead times of between one to three months. The strength of such lagged relationships indicate potential for developing useful climate-model-based forecasts at least one month ahead of time.

Following our exploratory analysis, our Bayesian spatiotemporal model framework (described in Section 4.1.2), in particular the specified DLNMs, allowed us to precisely quantify the directional relationship between DIRs and individual climatic variables. We estimated these relationships in the presence of other climatic variables and after accounting for seasonality effects, momentum effects, temporal random effects, and spatial random effects.

For each climatic variable, the three-dimensional exposure-lag-response relative risk (RR) plots (Figure 4.3) display, for different lags and different levels of a climate variable, the estimated elevated (or reduced) level of risk of incidence of new dengue cases relative to the risk induced by the corresponding mean value of the climatic variable. First, greater mean monthly maximum temperatures (above the mean value of 28°C) were associated with higher RR of incidence of new dengue cases, whilst lower values were associated with lower levels of RR. In terms of the cumulative risk induced by maximum temperatures (Figure SI B.9), higher levels of maximum temperature (e.g., 31°C) were consistently associated with greater cumulative risk, whilst lower maximum temperatures (e.g., 25°C) resulted in reduced cumulative risk. Second, the exposure-lag-response RR plot corresponding to total precipitation captures that extreme wet conditions in the preceding

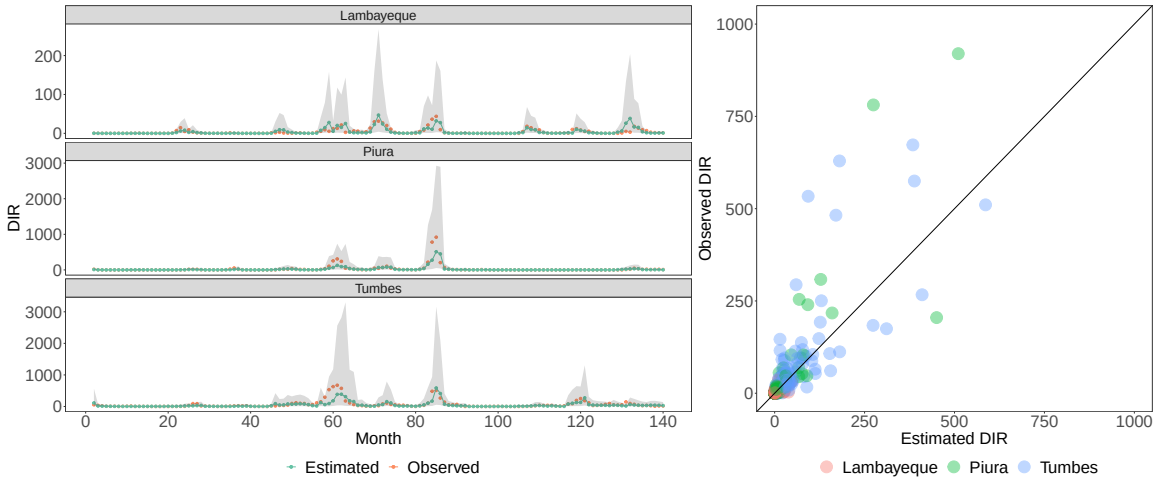


Figure 4.4: **Leave-one-time-point-out cross-validation predictive performance**. Left: The DIR (Dengue Incidence Rate) time series (gold) for each department is shown alongside the posterior median estimate (forest green) for each observation and the corresponding 95% credible intervals of the posterior predictive distributions (shaded grey). Estimates were obtained by refitting the model 139 times, excluding a single time-point/month (of three observations) one at a time, and estimating the posterior predictive distributions for the three omitted observations. Note that due to the presence of temporal autocorrelation terms (such as a random walk of order one prior distribution), leave-one-out predictive distributions for the observations at the first time-point are not generated. Right: The accompanying visualisation displays the observed DIR versus the corresponding estimated DIR, where the filled colours of light red, light green, and light blue represent Lambayeque, Piura, and Tumbes respectively.

four months were associated with elevated RR of DIR. Similarly, extreme low total precipitation values in the preceding months were found to contribute to higher RR, albeit at lower levels and only with a more substantial time lag of about one to three months. We further captured the precipitation findings in our plots of cumulative risk (Figure SI B.9), where extremities of total precipitation were associated with elevated cumulative risk, with the greatest risk occurring for extreme rainfall values over approximately three months. Third, the findings for SPI-6 were similar in terms of moderate-to-extreme rainfall and severe drought, albeit with a slightly different interpretation as SPI-6 is a cumulative and location-specific hydrological drought index which measures accumulated precipitation deficits over six months. Here, we found that lower values of this six-month-accumulated drought index (i.e. moderate-to-extreme drought), in any of the preceding one to four months, were associated with elevated RR and elevated cumulative risk, indicative of a lagged effect of drought conditions. Conversely, larger values of the SPI-6 (i.e. moderate-to-extreme precipitation) contributed to greater RR at more recent lags of zero to two months. Finally, in terms of the ICEN E-Index, we found that the greatest RR contribution, relative to the risk posed by the mean E-Index (-0.15), stemmed from IGP-classified *strong* El Niño events (greater than 1.7), followed by smaller RR contributions from *moderate* La Niña events (less than -1.4). In both cases, the lagged values of the index variable (by at least one month) were most important in terms of contributing to higher RR.

Climate-based retrospective modelling

Using our climate-based Bayesian spatiotemporal model, we accurately captured in-sample temporal trends in DIR from 2010 to 2021 across the three departments (Figure SI B.10), whilst also generalising well to synthetic out-of-sample environments. Performance of our model (described in Section 4.1.2) was appraised using a range of metrics such as DIC, CV log score, MAE, and cross-validated posterior predictive checking. The model's LOOCV posterior predictive check revealed that 94.7% of the DIR observations lay within the 95% credible intervals of our leave-one-out posterior predictive distributions. We display the out-of-sample predictive performance of our modelling framework in each department across the studied months in Figure 4.4, whilst model-based errors are visualised in Figure SI B.15. In general, there was a moderate-to-strong ability to track out-of-sample trends in DIR, ranging across substantial dengue outbreaks and periods of

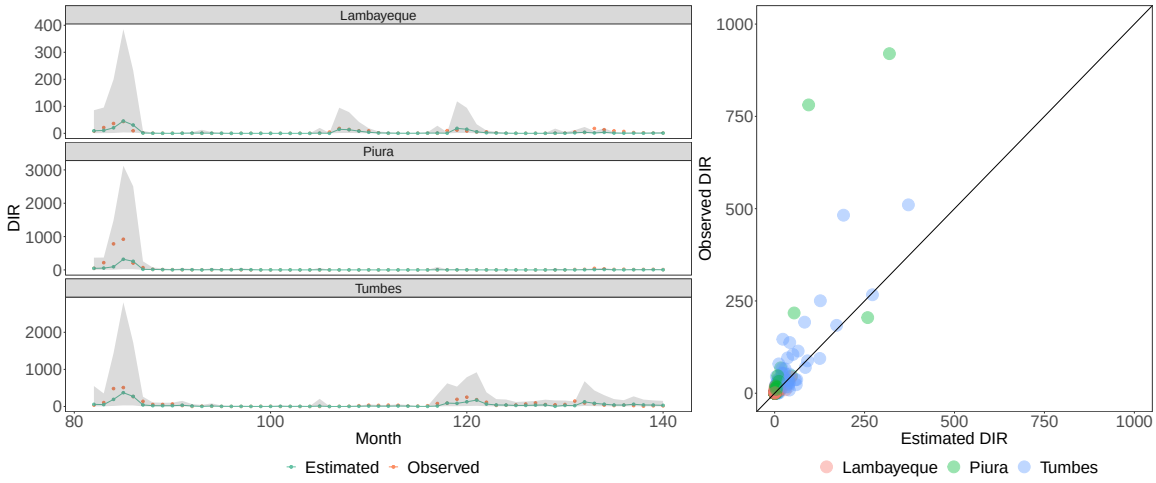


Figure 4.5: **Forecasting predictive performance across 2018 to 2021**. Left: The DIR (Dengue Incidence Rate) time series (gold) for each department is shown alongside the posterior median estimate (forest green) for each observation, and the corresponding 95% credible intervals of the posterior predictive distributions (shaded grey) which were obtained by fitting our model to the climatic and surveillance data up to one month preceding, and estimating the posterior predictive distributions for the next month's three observations. Right: The accompanying visualisation plots the observed DIR vs the corresponding estimated DIR, where the filled colours of light red, light green, and light blue represent Lambayeque, Piura, and Tumbes respectively.

lower DIR. We note a caveat that extreme large reported DIR observations (e.g., Piura in May 2018) occasionally resulted in under-estimation of the extent of an outbreak and widening of credible intervals (see Discussion).

To further assess our model's capabilities for reliable detection of outbreaks, we produced probabilistic estimates of out-of-sample observations exceeding (pre-defined) outbreak-indicative DIR thresholds of 50 per 100,000 and 150 per 100,000 (visualised, and stratified by true outbreak months and non-outbreak months, in Figures SI B.16 and SI B.17). In terms of out-of-sample outbreak classification (Table SI B.3), for outbreaks of DIR greater than 50 per 100,000 and greater than 150 per 100,000, our model enabled detection of 93% and 100% of such events respectively, and the respective false positive/alarm rates (the proportion of incorrectly classified non-outbreaks) were 0.13 and 0.06.

Climate-model-based forecasting

The analysis above presents our estimated climatic impacts on incidence of new dengue cases and retrospective results of our climate-based spatiotemporal model. Here, focusing on a public health use case, we appraise the forecasting utility of our model with a forecast horizon of one month. Across our forecasting analysis' studied period of 2018 to 2021, our model-based forecasting of the subsequent month's DIR was relatively accurate, as displayed by close correspondence between the observed DIR time series and forecasted values. Figure 4.5 visualises the association between the observed DIR values and the model-based forecasted values, whilst Figure SI B.18 captures the posterior median estimates of absolute errors. Specifically, 93.9% of DIR observations lay within the 95% credible intervals of the posterior predictive distributions. In general, large peaks in DIR values were correctly forecasted one month ahead, albeit with a caveat that extreme DIR observations in May to June of 2018 were under-estimated by posterior median estimates but lay within the relatively wide 95% credible intervals.

In terms of quantifying the anticipation of future outbreaks (Table 4.1, Figure 4.6) using our model (alongside historically calibrated outbreak cut-off probabilities), we correctly forecasted the occurrence of 100% of future dengue outbreaks where DIR exceeded 50 per 100,000 or 150 per 100,000. With respect to the reliability of the forecasts for outbreaks, false positive rates were 0.12 and 0.05 respectively, thus indicating relatively low probabilities of false alarms.

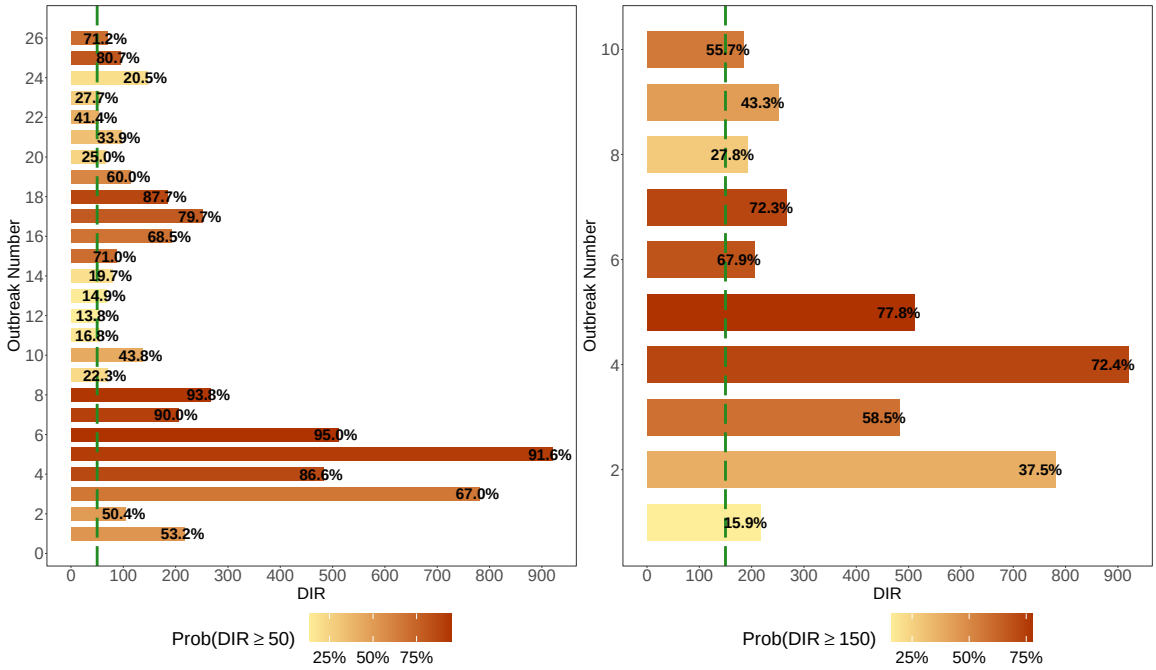


Figure 4.6: **Forecasting outbreak detection across 2018 to 2021.** Among the observations with DIR (Dengue Incidence Rate) greater than 50 (left) and DIR greater than 150 (right) per 100,000, the plots depict posterior probabilities of forecasted DIR exceeding thresholds (green) of 50 per 100,000 (left) and 150 per 100,000 (right). The plots capture the model's sensitivity in forecasting substantial dengue outbreaks one month in advance. To also visually assess the reliability of forecasted outbreaks and ensure a representative picture of our model's future outbreak detection capabilities, Figure SI B.19 is the analogous visualisation for the corresponding posterior probabilities of the observations with observed DIR less than the thresholds of 50 and 150.

4.1.4 Discussion

In our analysis, we developed a bespoke climate-based Bayesian hierarchical model which retrospectively modelled and prospectively forecasted spatiotemporal trends in dengue incidence across three Peruvian coastal departments. Retrospectively, across the twelve studied years from 2010 to 2021, we characterise our model's predictive performance by representative DIR estimates for individual observations that were excluded from the model-fitting process. Similarly, we produced reliable outbreak detection estimates in synthetic out-of-sample environments, including correct identification of 100% of out-of-sample outbreaks with DIR greater than 150 per 100,000, and a low false positive rate of 0.06. In our model-based analysis, which leveraged existing research[78,214] and our earlier exploratory analysis, we uncovered the following setting-specific inferences about the strength, timing, and structure of climatic influences on incidence of new dengue cases. Our model-based inferences were substantiated by extensive model checking across diagnostics and sensitivity analyses (which varied the model structure).

We found that greater maximum temperatures (between 29°C and 31°C) promoted heightened risk of incidence of new dengue cases, which is aligned with many clinical and observational studies of the *Ae. aegypti* vector and DENV. For instance, the optimal temperature for dengue transmission has been estimated to be approximately 29.3°C, whilst in Singapore, the overall cumulative risk (induced by maximum temperatures) for incidence of new dengue cases was estimated to have peaked at 31°C[82,497,498]. Here, our estimated elevated risk induced by greater maximum temperatures may be due to any of a myriad of factors including accelerated mosquito larvae development, faster reproduction rates, greater mosquito survival rates, reduced extrinsic incubation period of the dengue virus, heightened mosquito biting activity, and/or faster viral replication within the vector[80,81,217,392]. The estimated risk contributions from greater maximum temperatures at different time lags (across the four months) may be due to the time required for one or more of the following; temperatures to influence both larvae development and oviposition (with maximum temperature effects on oviposition activity in Ecuador

Table 4.1: **Summary statistics for forecasting outbreak months across 2018 to 2021.** DIR (Dengue Incidence Rate) observations (n = 180) were classified one month ahead of time as being predicted true outbreaks with DIR exceeding 50 (or 150) per 100,000 if the posterior probability of DIR exceeding 50 (or 150) was greater than a cut-off probability of 0.21 (or 0.13). *True Positive* measures the hit rate, or equivalently the proportion of true outbreaks correctly detected, whilst *False Positive* measures the proportion of non-outbreaks incorrectly classified as outbreaks. *Accuracy* represents the proportion of observations whose outbreak classification matched the true observed state (such as a predicted outbreak coinciding with an outbreak). *Precision* measures the proportion of predicted outbreaks which were true outbreaks. Finally, the *AUC* is the area under the Receiver Operating Characteristic (ROC) curve, is used as a measure of our model's skill for distinguishing between outbreaks. The cut-off probabilities were calibrated using historical data (prior to 2018) and thus, outbreak forecasting model performance is reflective of a realistic application of our framework by authorities. 95% CIs are the corresponding 95% confidence intervals[482,522]. Further analyses of outbreak classification performance were performed (see Table SI B.4).

Outbreak Threshold	True Positive	False Positive	Accuracy	Precision	AUC
DIR ≥ 50	1.00	0.12	0.90	0.59	0.94
(95% CI)	(0.87, 1.00)	(0.07, 0.18)	(0.85, 0.94)	(0.43, 0.74)	(0.77, 0.93)
DIR ≥ 150	1.00	0.05	0.95	0.53	0.97
95% CI	(0.69, 1.00)	(0.02, 0.10)	(0.91, 0.98)	(0.35, 0.87)	(0.96, 0.99)

estimated to be most significant at a six-week time lag), the mosquito life cycle (development occurs in four stages of egg, larva, pupa, and adult, which is estimated to be around two weeks), the incubation period of DENV (time between mosquito bite and the symptom onset of the infected individual, which is estimated to be three to ten days), and/or the time until dengue case notification[80,393].

We obtained similar findings for extreme precipitation events, especially heavy rainfall (reflected in our precipitation and SPI-6 variables). Heavy rainfall influences mosquito population dynamics by creating additional breeding habitats via the accumulation of stagnant water in settings such as water storage containers, discarded tires, and flower pots. Thus, ample breeding sites are created for *Ae. aegypti* to lay their eggs in stagnant water, and the mosquito population (in size and density) can increase, thus providing greater opportunity for mosquito-to-human contact. The effects of precipitation on the abundance of *Ae. aegypti* specifically are pronounced due to the preference of *Ae. aegypti* to oviposit in such artificial water storage containers (as opposed to vegetation) which can become feasible egg-laying sites due to precipitation[501,502,503]. Water storage practices in our studied departments, all of which are highly urbanised, have been influenced by extreme drought conditions and limits on daily water supplies which is amplified by the distinct dry and wet seasons. More generally, the role played by water storage containers in the generation of oviposition sites has been estimated to be dominant in dry seasons whereas solid waste (e.g., plastics) generally provide the sites in rainy seasons[523]. Our conclusion about the heightened risk posed by lagged extreme precipitation is in agreement with existing findings of stronger relationships for lagged values. For instance, aligning with our exploratory analysis, the strongest correlation with incidence of new dengue cases was estimated for two-month-lagged precipitation in Mexico and for one-to-two month lagged precipitation in Barbados, whilst in China, the strongest relationship was for lagged precipitation of three to four months[214,524,525].

Using our drought indicator (SPI-6), we estimated strong relationships between sustained drought conditions (accumulated precipitation deficits) and subsequent (lagged) heightened risk of incidence of new dengue cases. Our findings are similar to dengue research in other regions globally where similar lagged associations have been observed with drought periods[78,499,500]. The complex drought-induced risk may be due to changes to water storage and conservation patterns as households increase water storage in containers around the home, thus creating potential egg-laying sites for female *Ae. aegypti*[526,527]. As there may be fewer available breeding habitats, concentrated breeding may occur in the limited remaining water bodies. Here, lagged values of six-month-accumulated precipitation shortages were persistently associated with elevated relative risk of dengue transmission in the studied period, indicative of drought impacts manifesting themselves over a sustained period of several months. Combining the results for the extreme values of the SPI-6 (dry and wet conditions), we deduce a natural interplay between drought and subsequent rainfall. In particular, whilst drought conditions during

our studied period may have promoted favourable conditions for larvae habitat development (e.g., newly introduced water storage containers), habitat suitability can only be achieved in the presence of subsequent sufficient precipitation to enable the aquatic life cycle phase of the mosquito[528]. In arid climates such as Kenya, it is estimated that extreme flooding events have increased the egg and adult abundance[529]. The longer time lags between drought conditions and heightened dengue incidence risk may also be a result of the increased survival capacity of *Ae. aegypti* eggs in dry conditions (up to 120 days) and/or more gradual changes to water storage practices[78,84].

With respect to our model's final climatic covariate, associations between the ENSO and dengue outbreaks have been well-established for many years, as warming El Niño and cooling La Niña events can produce substantial anomalies in both precipitation and temperatures (which are conducive to mosquito development and dengue transmission)[499,530]. Here, we contributed to the existing evidence base surrounding the effects of ENSO on incidence of new dengue cases in the Peruvian coastal region. We uncover that the coastal-specific ICEN E-Index shared stronger associations than the ONI (which is measured for the entire Niño 3.4 region) and that large-magnitude values of the ICEN E-Index, lagged by one to four months, induced elevated risk of incidence of new dengue cases[514,531]. We identified that; i) larger positive values of the ICEN E-Index (reflective of strong El Niño events), at lags beyond two months, produced higher risk of incidence of new dengue cases, possibly due to warmer El Niño-induced ambient temperatures and drought conditions (although heavy rainfall events can also be induced by El Niño), and ii) more extreme negative values of the E-Index in the preceding one to two months (reflective of La Niña events) were also associated with greater risk of incidence of new dengue cases (albeit with smaller absolute effects), presumably due to the resulting extremely wet conditions. The findings display that lagged effects are particularly important for modelling incidence of new dengue cases in the coastal region, as SST anomalies naturally take time to culminate in extreme meteorological events onshore. These discoveries also reinforce our finding that risk induced by drought conditions (typically from El Niño events) operated at longer time lags than extreme rainfall (typically from La Niña events), and when restricting ourselves to the Peruvian coast, El Niño events are more pertinent (compared to La Niña events) for modelling incidence of new dengue cases.

Our analysis is subject to limitations. First, the reliability of our inferences is dependent on the quality of the climate and surveillance data used. For example, due to the dengue surveillance data available for Peru, we cannot directly account for individual serotypes of DENV (as two serotypes are known to circulate in the studied departments; DENV-1 and DENV-2) or variable reporting rates (both in time and space). The reliability of surveillance data may vary during the studied period. For instance, under-reporting may have resulted from events such as the damage caused to 299 health care centres in Piura during the 2016-2017 ENSO, which could have been additionally confounded by increased notification of other infectious diseases (such as Zika and leptospirosis)[532]. Such variability in reporting rates within and across regions and years may have impacted the model-fitting process, and yielded substantial overdispersion and/or artificially affected accuracy measures. We cannot exclude the possibility of additional unobserved confounders in the form of socioeconomic conditions such as population density, clean water supply, living conditions, education level, sanitation infrastructure, and access to healthcare[393,533,534,535]. For example, poor access to water supplies promotes storage in potable water containers, particularly during drought conditions, thus stimulating mosquito abundance[536]. In contrast, access to private water wells and health education status are negatively associated with indices of larval abundance[537]. Similarly, we do not directly estimate the time- or space-varying risks induced by climatic influences. Thus, we indirectly adjusted for such unobserved confounders in our Bayesian spatiotemporal model (Section 4.1.2) via unstructured spatial effects, whilst we included yearly random effects to adjust for plausible year-to-year heterogeneities (such as changes in reporting rates or vector control programmes).

Alongside our investigation of climatic influences over time, we have presented an important use case for our climate-based model framework; reliable, robust forecasting of a subsequent month's incidence of new dengue cases. For forecasting of outbreak events, we report strong forecasting skill, including detection of 100% of all outbreaks with DIR $\geq$ 50 (or 150)

per 100,000 one month ahead of time, and a corresponding AUC of 0.94 (0.97). We note two caveats and thus, potential for future model improvements; i) DIR observations with values marginally above the defined outbreak thresholds possessed low forecasted model-based posterior probabilities of breaching thresholds, and ii) due to the dependence on autocorrelation effects and a momentum oscillator, our model can currently only forecast incidence of new dengue cases with one month forecast horizons. Future work may aim to develop and implement a forecasting model with greater forecast horizons, include additional socioeconomic variables and variability in climatic conditions (e.g., diurnal temperature range), and/or refocus on DIR at a district or province level (though higher resolution data could present additional challenges e.g., spatial heterogeneities in reporting rates).

Nevertheless, using our current modelling framework, we produced representative estimates of future trends in new dengue cases and reliable classifications of future outbreaks (with robust detection of future public health crises and emergencies). Thus, the modelling framework and associated forecasts could be readily integrated into future early warning systems in Peru, and assist in the well-timed deployment of vector control resources and information campaigns. Surveillance and response strategies are coordinated by the Ministry of Health in partnership with regional (coarser spatial resolution than departments) health authorities [504]. Thus, existing strategies (e.g., targeted larvicidal treatments of standing water and insecticide spraying) could be enhanced in their timing, effectiveness, and efficiency. Due to low probabilities of false alarms, public health authorities could have a high degree of confidence in the extensive capital and resource investment involved in such vector control campaigns. Such targeted campaigns could help to control the size of future outbreaks, reduce healthcare pressures, and potentially lower dengue-related morbidity and mortality. Hence, we conclude that contingent on the timely availability of high-quality surveillance data and climate data, authorities could employ our interpretable, climate-based modelling framework to i) model and forecast current and future trends respectively in incidence of new dengue cases and ii) quantitatively inform public health decision-making.

4.2 Paper II: Multi-model approach to understand and predict past and future dengue epidemic dynamics

Cathal Mills, Francesca Falconi-Agapito, Jean-Paul Carrera, César V. Munayco, Moritz U. G. Kraemer, and Christl A. Donnelly. Multi-model approach to understand and predict past and future dengue epidemic dynamics, June 2025. URL https://www.medrxiv.org/content/10.1101/2024.10.18.24315690v2. Pages: 2024.10.18.24315690

In this follow-up manuscript, currently under review at *Royal Society Open Science*[46], we extended the methods above to develop a multi-model workflow for modelling past dynamics and drivers and forecasting future dengue incidence at finer spatial resolutions.

Abstract

Understanding the past, current, and future dynamics of dengue epidemics is challenging yet increasingly important for global public health. Using data from northern Peru across 2010 – 2021, we introduce a multi-model approach that integrates new and existing techniques for understanding and predicting dengue epidemics. Using wavelet analyses, we unveil spatiotemporal patterns and estimate space-varying epidemic drivers across shorter and longer dengue cycles, while our Bayesian hierarchical model allows us to quantify the timing, structure, and intensity of such climatic influences. For forecasting, as a single model is generally sub-optimal, we introduce trained and untrained probabilistic ensembles. In settings that mirror real-world implementations, we develop climate-informed and covariate-free deep learning forecasting models involving foundational time series, temporal convolutional networks, and conformal inference. We complement modern techniques with statistically principled training, assessment, and benchmarking of ensembles, alongside interpretable metrics for outbreak detection to disseminate outputs with communities and public health authorities. Our ensembles generally outperformed individual models across space and time. Looking forward, whether the public health objective is to learn from the past and/or to predict future dengue epidemic dynamics, our multi-model approach can be used to inform the decision-making of public health authorities.

Author Contributions

- Conceptualization: **C.M**.
- Methodology: **C.M**; C.A.D; M.U.G.K
- Data curation: **C.M**.
- Data visualisation: **C.M**.
- Writing original draft: **C.M**; C.A.D; M.U.G.K.
- Writing final draft: **C.M**; C.A.D; M.U.G.K.; F.F.A; J.P.C; C.V.M

4.2.1 Introduction

Dengue is a mosquito-borne disease which poses an increasingly global public health threat. Reported annual incidence has risen from 500,000 cases in 2000 to over 14 million cases in 2024[41]. 80% of the reported cases in 2023 occurred in the WHO Region of the Americas, yet dengue is endemic in over 100 nations and is expanding in distribution within and across nations[67,72,355]. Primarily spread by *Aedes* mosquito vectors (female *Ae. aegypti* and *Ae. albopictus*), the viral

infection is caused by the dengue virus (DENV), an orthoflavivirus with four genetically distinct serotypes and multiple lineages[64]. Infection with an individual serotype offers long-term immunity to the specific serotype, but only short-term immunity to others[538]. Symptoms of dengue vary in severity, and up to 80% of all infections are asymptomatic[67].

The increasing number of dengue outbreaks has been linked to human, viral, and environmental factors. Climatic conditions such as temperature, precipitation, and humidity influence the mosquito abundance and DENV replication, as *Aedes* mosquitoes thrive in warm, humid conditions and precipitation is necessary for a mosquito's juvenile stages[81,497,501,503]. Extreme climatic conditions, including drought and heavy precipitation, instigated by events such as the El Niño phenomena, have been identified as epidemic drivers[45,352,531,539]. Meanwhile, modulating human factors like rapid urbanisation have allowed for the creation of mosquito breeding sites (e.g., water storage practices) and addition of susceptible hosts in close proximity, while increased travel, human mobility, and globalisation have similarly contributed to greater potential for rapid geographic spread[77,85].

Faced with a difficult-to-detect, climate-sensitive disease, public health authorities require a quantitative understanding of epidemic dynamics and drivers to inform their decision-making. In the absence of complete spatial, temporal, or population coverage of dengue epidemiological surveillance, authorities may also need alternative ways to fill gaps in surveillance[540]. Prospectively, public health authorities often aim to have early warning systems and predictive models of spread which can probabilistically predict future epidemic trajectories and the likelihood of outbreaks, thus allowing sufficient time for design of evidence-based public health policies[159,246,541,542]. Therefore, in Brazil, South-east Asia, and Barbados, modellers have used wavelets, Bayesian hierarchical models, deep learning models, and ensemble frameworks to understand epidemic dynamics and forecast incidence and outbreaks. These approaches have separately enabled a model-based understanding of spatial patterns (e.g., highly correlated outbreaks and amplifying effects), temporal patterns (e.g., strong seasonality), demographic patterns (e.g., mean ages of infection), and key epidemic drivers (e.g., climatic and human influences)[78,168,214,438,539,543]. In our studied region of northern Peru, dengue is a substantial, year-round public health burden where similar techniques have been used (separately) to quantify climatic effects and forecast short-term epidemic trajectories[45,531,544].

Here, we build upon published work, introducing an interdisciplinary modelling workflow (Figure 4.7). To date, modelling workflows have been presented for a single modelling discipline, yet are not dengue-specific nor enable evidence synthesis from a multi-model approach[188,457,545,546]. Workflows also rarely integrate a multi-model approach that can learn from the past, monitor current, and predict future dynamics. As a model is an imperfect description of reality, there is no reason to assume (potentially restrictively) that a single technique for modelling is correct and appropriate for decision-making[105,160]. Here, we acknowledge this fact to avoid drawing results and conclusions from a single modelling discipline, as our approach leverages strengths from statistical time series, wavelet analysis, Bayesian statistics, deep learning, and probabilistic ensemble forecasting. We demonstrate how our approach might be used in real-time to robustly analyse past, current, and future dengue epidemic dynamics. Using fourteen provinces in Peru as a case study we illustrate the potential utility of our approach, comparing it to more established dengue models by specifically focusing on the advances brought by integrating multiple techniques in terms of accuracy, reliability, and robustness. Our modelling outputs and general framework are generalisable and can be used across any context where dengue outbreaks occur frequently (endemic settings).

Epidemiological modelling for public health policy can be viewed through the lens of three axes; explanation, prediction, and decision analysis, and our framework enables explanatory and predictive insights from seemingly distinct approaches that aid decision-makers with monitoring and controlling dengue transmission. Thus, whether the objective is to understand (modelling climatic drivers) or predict (forecasting dengue cases), our framework enables robust conclusions and thus bridges some of the gaps in the literature regarding modelling and forecasting beyond single domains.

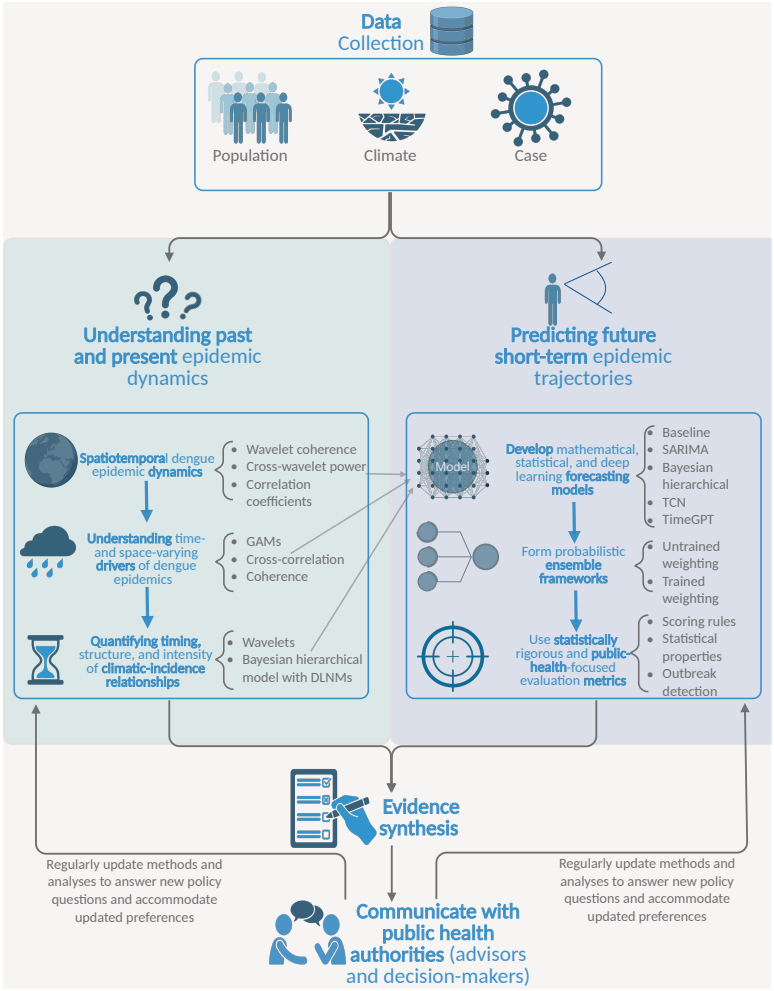


Figure 4.7: **Modelling and forecasting workflow.** Our multi-model workflow takes the researcher and public health authority from data collection to studying past, present, and future dengue epidemic dynamics. The workflow begins by sourcing heterogeneous data types. Our retrospective analyses use wavelets, Bayesian modelling, and other techniques to provide a quantitative understanding of past epidemic dynamics and drivers. This pillar (left) focuses on understanding dengue epidemics and may be traditionally regarded as exploratory analysis, yet we avoid this distinction as wavelets are not a model in the traditional sense yet are still a mathematical tool that requires specification of the type of wavelet and other hyperparameters. For forecasting future dynamics, we propose multi-model probabilistic forecast ensembles which predict new dengue cases with a horizon of up to one month, while our forecast evaluation involves both statistically rigorous (e.g., proper scoring rules) and public-health-focused criteria (e.g., outbreak detection metrics). The outputs of each stage of the retrospective analyses (left pillar) can inform the development of individual forecasting models. From the multi-model retrospective and forecasting analyses, evidence synthesis across space and time is a vital component for operationalising a useful workflow and for communicating the modelling outputs with different stakeholders in public health authorities (e.g., scientific advisors and decision-makers may require different levels of modelling details). See Sections 4.2.3 and 4.2.4 for an illustrative application and evidence synthesis for the workflow.

4.2.2 Materials and methods

Data

We used province-level reported dengue cases per month from May 2010 to December 2021 across the fourteen provinces in Piura, Tumbes, and Lambayeque (on the northern coast of Peru) which had registered large dengue outbreaks during this period, as per the National Centre for Epidemiology, Disease Prevention and Control (Peru CDC) in Peru's Ministry of Health[505]. The reported cases included both confirmed and probable cases across all serotypes (without serotype classification for each case). Probable dengue cases were defined as an individual with i) febrile illness for a maximum of

seven days, ii) two or more specific symptoms, and iii) resides in or has recently visited areas with dengue transmission or known *Ae. aegypti* populations. Confirmed dengue cases met the same exposure and symptoms criteria, and had a positive dengue test[504]. We aggregated weekly case data (using epidemiological weeks, starting from Sunday and ending on Saturday) to total monthly cases to align with other data sources (see below). In real-time applications of our framework, it is unclear the extent of delay (potentially varying across time and space) in reporting individual dengue cases in Peru, yet we suspect that the delay is 1–2 weeks and varies between health regions. Such a phenomenon which could impact the timeliness and utility of real-time modelling and forecasting analyses.

We interpolated population data from the 2007 and 2017 national censuses, and the 2022 mid-year population estimates[547,548,549]. Then, we defined Dengue Incidence Rates (DIRs) per 100,000 by converting reported new dengue cases to rates per 100,000. Data on the proportion of provinces' populations living in urban areas were also sourced from the 2007 and 2017 censuses but was not available more recently. We therefore used the 2017 data for years up to 2021.

Climatological data were as previously described in[45] and we calculated average values for each province area. We sourced total monthly precipitation and monthly averages of daily maximum (and minimum) temperature from the WorldClim 2.1 dataset[281,509], a drought indicator called the Standardised Precipitation Index (SPI-6)[510,511] from the European Drought Observatory[512], and El Niño Southern Oscillation (ENSO) indicators, namely the El Niño Coastal Index (ICEN) and the Oceanic Niño Index (ONI) from the Geophysical Institute of Peru (IGP)[513,514] and the National Oceanic and Atmospheric Administration[515] respectively. Each of these climatological variables has been used in previous modelling studies in Peru and/or in other nations[45,78,79,214,531].

Wavelet and exploratory analyses

We developed wavelet methods to analyse epidemic dynamics and drivers. Wavelets have been used for dengue analyses in neighbouring Brazil and across Southeast Asia[438,539,543], and can analyse signals with sharp discontinuities and time-varying periodicity (i.e. non-stationarity). Using the R package WaveletComp, we used a Morlet wavelet with non-dimensional frequency $\omega_0 = 6$[495,539,543,550], which allowed us to decompose DIRs into reconstructed annual (maximum period of two years) and multiannual (period of two to twelve years) cycles. The wavelet transform produced a wavelet power spectrum, and we averaged across the provinces' reconstructed annual and multiannual cycles to obtain the average wavelet power of the annual and multiannual cycles per province per month.

For both annual and multiannual cycles, to compare the relative timing of epidemics for each province pair, we computed the cross-wavelet power (the wavelet analogue of covariance) and wavelet coherence (the wavelet analogue of correlation). To measure epidemic synchrony (how the amplitudes of the incidence time series covary), for each province pair, we computed the pairwise Pearson correlation coefficient between their DIR time series and between their reconstructed (annual and multiannual) cycles. We also quantified time- and space-varying geographic, human, and climatic drivers via the cross-wavelet power and coherence between climatic conditions and reconstructed dengue cycles. We complemented this by developing generalised additive models (GAMs) for epidemic synchrony using smooth functions (thin-plate splines) of temporal random effects (month and year), spatial random effects (province), climatic variables (all variables above), and human mobility approximations (the pairwise product of provinces' populations)[551,552]. We fitted GAMs using the mgcv package in R version 4.2.1[495,552], and evaluated model performance using the minimised generalised cross-validation score (GCV), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)[464,465,553].

Other exploratory analyses involved computing, for each province, the cross-correlation between past climatic conditions and current DIR at different time lags.

Bayesian climate-based modelling

We developed a Bayesian climate-based modelling framework which was similar to existing models from other nations[78,79,214] and our previous work in Peru[45]. Here, we focus on recent modifications.

We modelled province-level dengue incidence using a zero-inflated Poisson distribution, thus accounting for anticipated excess zero counts. The final model, informed by exploratory analyses, was developed using data across 2010 to 2017, and contained: i) fixed effects (to estimate effects of recent incidence trends relative to the preceding year, shared seasonality across provinces, urbanised populations, and epidemic momentum), ii) province-level temporal effects (to estimate their monthly patterns and year-to-year heterogeneities), iii) spatiotemporal random effects via an adapted Bayesian version of the Besag-York-Mollié model[519,520] (to account for similarities and uniqueness across provinces), and iv) nonlinear, delayed climatic influences via distributed lag nonlinear models (DLNMs)[394]. We provide additional model details in Appendix C.4.

We assessed models using information criteria (Widely Applicable Information Criteria, WAIC and cross-validated logarithmic score), in-sample predictive accuracy, and leave-one-time-point-out cross-validation (alongside central prediction interval (PI) coverage and probabilistic calibration). We fitted the models using Integrated Nested Laplace Approximation (INLA)[459,495].

Probabilistic ensemble forecasting

By sampling from the posterior predictive distribution, the climate-based model directly enables probabilistic forecasts of dengue incidence with a forecast horizon of up to one month. However, individual models can have contrasting strengths and be wrong in different ways (e.g., predictive vs understanding transmission mechanisms). So, we consider ensemble frameworks which combine the forecasts of individual models. Ensemble frameworks have demonstrated strong predictive performance for many dynamical systems (e.g., COVID-19, seasonal influenza, dengue, and weather forecasting[162,169,172,554]). Ensemble forecasts are often represented in one of two ways; using samples or quantiles of the predictive distribution. We focus on quantile-based forecasts as i) it provides access to a wider range of models which directly target quantiles (e.g., where there is no assumed likelihood), and ii) it is a computationally cheaper environment (versus processing thousands of samples). These have been used in recent years[58,433]. Some ensemble approaches for dengue have not used probabilistic forecasts (e.g.,[555,556]), instead focusing on pointwise accuracy metrics which do not measure uncertainty in our predictions. These approaches also have not always separated the model development period from the testing period nor benchmarked performance using iteratively updating models, all of which is necessary to reflect a realistic application. In recent years, probabilistic ensembles have been increasingly used, often with equal weighting or trained weighting of sample distributions[162,169,554,557].

For the forecasts of dengue incidence at a one-month horizon, we used 23 quantile levels (identical to past forecasting studies[433]). For untrained approaches, we used the median or mean average of individual modelling components' predictive quantiles, while for trained approaches, we used linear stacking with the quantgen package in R[558] which weighted individual models' quantile forecasts for a given month based on weights that optimised the sum of the quantile losses for forecasts over the past twelve months. We performed the trained weighting for each province independently (i.e. province-dependent weights) and for all provinces jointly (i.e. province-independent weights). We iteratively updated the model weights as each month's data became available (i.e. a realistic online learning procedure).

While we performed our study retrospectively, to reflect a realistic application, we forecasted dengue cases with a one-month horizon using an expanding window which left all future data out of model training. The four-year period of January 2018 to December 2021 was our testing period. This allowed sufficient data for the model development period (2010 to 2017 inclusive), where we tested forecasting models with leave-future-out cross-validation (also known as time series cross-validation[559]. We selected covariates of each model (Section 4.2.2) solely from performance of candidate

models in the model development period.

To evaluate forecasts, using the scoringutils R package[560], we applied the weighted interval score (WIS) – a proper scoring rule that approximates the continuous ranked probability score (CRPS)[58,470]. The WIS measures the absolute distance between the forecast/predictive distribution and observed data, and simultaneously assesses model calibration (compatibility of forecasts with observations) and sharpness (concentration of predictive distributions). Lower scores indicate better performance. As the WIS accounts for distributional uncertainty, we avoided focusing excessively on pointwise accuracy metrics. It matters to our forecast assessment how we apply a scoring rule, and applying to case counts is likely inappropriate due to different epidemic properties across space and time[213]. Here, we i) follow recommended procedures to score forecasts on a natural-logarithm-transformed scale (i.e. log(Cases + 1)), and ii) also score forecasts on a DIR scale. While we primarily focused on the recommended logarithmic scale, the latter assessment is motivated by a desire to be comprehensive and to provide interpretable, comparable information for public health authorities across space and over time. Alongside the WIS, we assessed bias, quantile coverage, and PI coverage, which measure respectively whether the model is systematically wrong in a given direction, whether the model's quantile predictions are well-calibrated and whether the model's prediction intervals (PIs) are well-calibrated. These statistical properties quantify reliability of forecasting models, and may guide forecasters and public health authorities when deciding whether a model can be employed for consistently, stable results.

While all such measures provided aggregate and statistically rigorous perspectives of performance, we also included a public-health-focused measure which focused attention on times of the year and areas of the predictive distribution of public health urgency. Here, we evaluated ability to correctly classify (i.e. binary forecast) the following month's DIR reaching thresholds of 50 or 150 per 100,000. These thresholds were pre-specified by us to indicate different magnitudes of dengue outbreaks, and we report classifications which can be easily interpreted by public health authorities and the wider public. We used our model-based quantile forecasts to identify a cut-off quantile level (i.e. a decision rule based on historical performance) which determined whether we classify an outbreak (of DIR $\geq$ 50 or 150 per 100,000) for the next month. An outbreak is forecasted if the prediction at the (historically calibrated) cut-off quantile level exceeded the corresponding outbreak threshold (e.g., 50 per 100,000). We allowed decision rules to be iteratively updated as each month's data became available – in each month, the rule was determined to maximise the historical AUC (the area under the receiver operating characteristic curve) for classifying outbreaks. This separated the inference stage (modelling) from the decision-making stage, and ensured that our classifications maximised public health utility (simultaneously maximising true positives and minimising false alarms avoids wasted resources and lack of public confidence). We repeated this outbreak detection evaluation for the first onsets of outbreaks, where we assessed performance for forecasts of initial yearly onset of outbreak with DIR $\geq$ 50 (or 150) per 100,000. This smaller dataset includes classifications for each province for months of a year corresponding to the outbreak onset, months before the onset, or in a year with no eventual onset. This classification focuses on a public health priority of dengue incidence *first* reaching pre-defined thresholds.

Our ensemble framework components

Our ensemble frameworks contained two or more forecasting model components whose predictive distributions (for one-month-ahead cases) were represented by predictive quantiles (23 levels as used in previous forecasting studies[170,172]).

Our baseline model was an epidemiologically naive model which forecasted next month's dengue cases as the current month's dengue cases. This was an application of an existing benchmarking model from the U.S. COVID-19 Forecast Hub, and is a random walk model with innovations (i.e. uncertainty) based on past differences in monthly dengue cases[433].

Our Bayesian climate-based model (described in Section 4.2.2) used information up to one month before the target month to generate samples from a posterior predictive distribution – from which we computed quantile forecasts. We also developed Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models. These statistical time series models

Table 4.2: **Quantile-based forecasting models and abbreviations.** The six individual forecasting models are described first (above the double lines), followed by ensemble models that are highlighted with an asterisk. The equal weighting (EW) scheme applies equal weights to the quantile of that ensemble model's individual predictions for each quantile. The median of individual models means that we take the median of the individual models' predictions for each quantile.

Model Abbreviation	Description
Baseline	Epidemiologically naive random walk model that uses the most recent month's observed cases as a forecast for the next month.
Bayes-Climate	Bayesian spatiotemporal model with nonlinear, delayed climatic effects, spatial random effects, a momentum indicator, and temporal effects
SARIMA	Seasonal Auto-Regressive Integrated Moving Average model specified for each province independently at each time
TCN	Temporal Convolutional Network with climatic covariates specified for each province independently
TimeGPT	Transformer-based foundational time series forecasting model with climatic covariates specified for each province independently
TimeGPT-NoCov	Transformer-based foundational time series forecasting model without any covariates specified for each province independently
EW-Mean *	Equally weighted mean average of all individual forecasting models
EW-Mean-NoBase *	Equally weighted mean average of all individual forecasting models, excluding the baseline model
EW-Mean-NoBayes *	Equally weighted mean average of all individual forecasting models, excluding the Bayesian climate-based model
EW-Mean-NoCov *	Equally weighted mean average of all covariate-free individual forecasting models, excluding covariate-based models (the Bayesian climate-based, TimeGPT, and TCN models)
Median *	Median of all individual forecasting models, including the baseline model
Median-NoBase	Median of all individual forecasting models, excluding the baseline model
Median-NoBayes *	Median of individual forecasting models, excluding the Bayesian climate-based model
Median-NoCov *	Median of individual forecasting models, excluding covariate-based models (the Bayesian climate-based, TimeGPT, and TCN models)
Prov-Trained*	Trained ensemble with province-dependent weights determined to minimise the sum of the quantile losses
Trained*	Trained ensemble with spatially homogeneous weights determined to minimise the sum of the quantile losses

were fitted independently to the logarithm of monthly cases for each province independently using the Auto-ARIMA (Auto-Regressive Integrated Moving Average) implementation of the Statsforecasts package and Darts library in Python[561,562]. This implementation selects the order of the ARIMA model based on the AIC[464], and quantile forecasts are generated by assuming normality in forecast errors. We refitted the model at each time point to reflect real-world conditions as new information becomes available. For the baseline, Bayesian, and SARIMA models, we generated quantile forecasts by taking quantiles of the 5,000 samples of the predictive distributions.

The models above are traditional model frameworks, yet in recent years, deep learning approaches have demonstrated strong forecasting performance for infectious diseases[397]. Informed by our exploratory analyses, we introduce here the first known application of a foundational time series model for forecasting dengue cases. We used TimeGPT, a transformer-based foundational time series model with self-attention that was pre-trained on over 100 billion data points from different domains[401,563]. We implemented this model (for the logarithm of cases) with and without climatic covariates (two-month rolling averages of precipitation and minimum temperature). The model readily produces quantile forecasts by using conformal inference — a distribution-free way to produce statistically rigorous PIs[426]. We also developed a forecasting model that used a temporal convolutional network (TCN) from the Darts library in Python[561]. This deep learning architecture has dilated causal convolutional neural networks which preserve the local and temporal structures of the data. To avoid making distributional assumptions, we used quantile regression to directly estimate the quantiles of the predictive

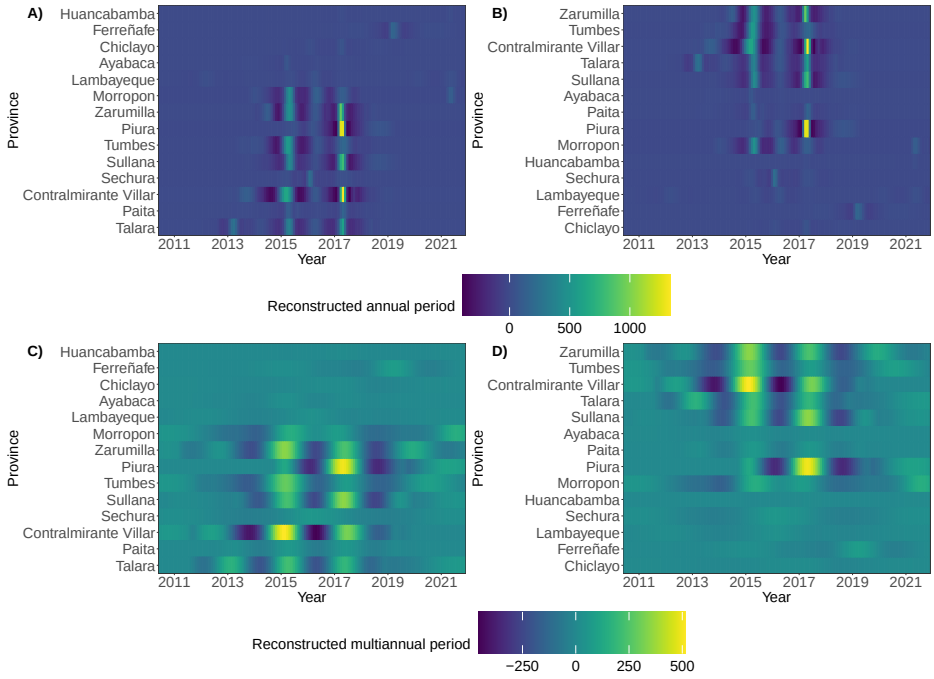


Figure 4.8: **Wavelet reconstructions of cycles of Dengue Incidence Rate (DIR) time series. (A)–(B)**: Reconstructed annual cycles of province-level DIRs, where provinces are sorted in ascending order by **(A)** longitude (from west to east) and **(B)** latitude (from south to north). **(C)–(D)**: Reconstructed multiannual cycles of province-level DIRs, where provinces are again sorted in ascending order by longitude and latitude respectively. Higher values of the annual (multiannual) period indicate peaks in DIR across the annual (multiannual) reconstructions, where the wavelet-based reconstructions are the reconstructions of province-level DIR time series using the initial decompositions into time-frequency and then, frequencies corresponding to annual (multiannual) cycles. Figure SI C.2 depicts the reconstructed cycles with provinces sorted by percentage of population living in urban areas.

distribution. We again used climatic covariates of two-month rolling averages of precipitation and minimum temperature. Covariates were based on exploratory analyses from the model development period.

We outline model abbreviations in Table 4.2 and additional details on ensemble forecasting models in Appendix C.5. We followed EPIFORGE 2020 guidelines[564] throughout the forecasting analysis (Table SI G.2).

4.2.3 Results

Investigation of epidemic dynamics using wavelets and climate-based analyses

We used wavelet analysis to simultaneously analyse signals (of incidence and climatic time series) in both time and frequency domains, focusing on reconstructions of annual and multiannual (two to five years) cycles of DIRs. Similar to dengue analyses from Southeast Asia[438,539], large dengue epidemic years (2015 and 2017) were characterised by statistically significant peaks in the average wavelet power and amplitude of reconstructed cycles (Figures 4.8 and SI C.3). Spatially, we found common large peaks across reconstructed cycles in provinces of the north and west, with slightly more widespread incidence peaks in the second large epidemic year of 2017. The mean period of multiannual cycles reduced during such years, while across the study period, we observed similar overall trajectories of the mean period (Figure SI C.7) across the provinces (median pairwise Pearson correlation: 0.66, Interquartile Range – IQR: 0.25, 0.84). Complementary visualisations of monthly DIRs suggested strong similarity in the province-level dengue incidence, in terms of both timing and magnitude. Trends in dengue incidence included waves in the south preceding increases in provinces in the north and waves in the west preceding surges in the east. The trends were observed within individual years and when analysed using monthly averages over the study period (Figures SI C.4 – SI C.6).

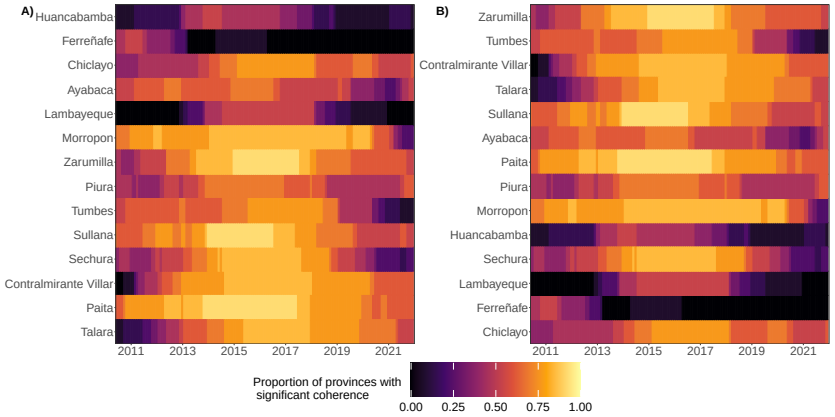


Figure 4.9: **Coherence of reconstructed annual dengue cycles.** Coherence is the wavelet analogue of correlation and here, compares the relative timing of provinces' dengue epidemics. We visualise the proportion of a focal province's annual dengue cycles which share statistically significant coherence with the cycles of other provinces at each time point. Provinces are sorted in ascending order by **(A)** longitude (from west to east) and **(B)** latitude (from south to north).

Wavelet coherence (measuring the relative timing of epidemics) was generally widespread across time and space (Table SI C.5, Figures SI C.11, SI C.10). Coherence amplified (i.e. was more widespread) during epidemic years when more provinces shared statistically significant coherence with the (annual and multiannual) dengue cycles of other provinces (Figure 4.9). Spatially, more northerly and more westerly provinces tended to possess a greater proportion of statistically significant coherent relationships with annual and multiannual cycles of other provinces, whilst more southerly and easterly provinces had a larger number of out-of-sync (or equivalently, self-contained) cycles. We estimated these spatial trends at individual time points within years (Figure 4.9) and across the studied period of 140 months (Figure SI C.11), yet a caveat is that Chiclayo (the most southerly province) had more coherent cycles, reflective of more urbanised provinces possessing more significantly coherent annual and multiannual cycles (Table SI C.6, Figure SI C.2).

We found moderate-to-strong epidemic synchrony between annual raw DIR time series. Similar to coherence, there was strong synchrony between reconstructed cycles (both annual and multiannual, Table SI C.7), and synchrony was high during large epidemic years (Figure SI C.13). Spatially, particularly in reconstructed multiannual cycles, synchrony tended to be stronger in more northerly and westerly provinces (Figure SI C.12), which indicates that correlations between epidemic curves shared similar patterns as the seasonality of epidemics (i.e. coherence).

Using coherence and cross-wavelet power, we quantified impacts of geographic, climatic, and human drivers on reconstructed cycles, and found widespread statistically significant coherence and consistently greater average cross-wavelet power between climatic variables (maximum temperature, precipitation, and ICEN) and dengue cycles (both annual and multiannual) during epidemic years of elevated incidence (Figure SI C.14). Peaks in average cross-wavelet power were always observed during periods of elevated incidence across each of the climatic variables and across almost all of the provinces (Figures SI C.15, SI C.16, SI C.17). More northerly provinces shared greater levels of statistically significant coherence with climatic variables. The ICEN had a greater coherence (with reconstructed cycles) in more westerly (i.e. coastal) provinces, and had a spatial gradient over time in the cross-wavelet power with annual cycles, as we estimated greater agreement with seasonality of cycles over time as one progresses from west to east and from north to south (Figure SI C.17).

To assess the influence of lagged climatic conditions on human disease incidence and to inform model development, we computed cross-correlations between climatic variables and the raw monthly DIR time series (during the model development period; 2010 to 2017). There were generally moderate-to-strong, and spatially consistent, relationships between climatic conditions and DIRs across the provinces, particularly across temperature, drought, and precipitation. From a forecasting

perspective, the strongest relationships were inferred with climatic lead times of one to three months, and cyclical patterns in the relationship were visibly pronounced (Figure SI C.1).

We investigated the impact of pairwise distance between provinces on their corresponding average correlation and proportion of significant coherent cycles. Within some provinces, there were clear relationships between shorter distances and greater average correlation or coherence (which was most widespread for coherence between multiannual cycles). Shorter pairwise province distances were associated with similar annual epidemic dynamics (timing and amplitudes, across annual and multiannual cycles) in more northerly provinces (Figures SI C.18 and SI C.19).

We extended previous modelling methods[438] (also incorporating the pairwise product of population sizes) as we developed GAMs to identify predictors of epidemic synchrony (see Materials and methods). Our best-fitting GAM, based on generalised cross-validation score and corrected AIC[565,566], included climatic influences (minimum temperature and precipitation), year-to-year heterogeneity, and the (logarithmic) pairwise product of population sizes. Here, positive (negative) coefficients indicate estimated enhanced (diminished) synchrony in the presence of the corresponding covariate. We estimated negative coefficients in twelve of the fourteen provinces for the pairwise distance-province interaction effect on epidemic synchrony, while six of the fourteen provinces had statistically significant (all negative) effects (Figure SI C.20). Significant positive (partial) effects were also estimated for temporal (yearly) random effects during epidemic years, whilst lower (and higher) values of the pairwise product of populations were associated with statistically significant negative (and positive) effects on synchrony respectively. Climatic partial effects were subject to greater uncertainty, but greater minimum temperatures and greater precipitation were estimated to have increasingly negative and increasingly positive effects respectively (Figure SI C.20 and see Discussion).

Bayesian climate-based modelling

The results from our climate-based Bayesian spatiotemporal model complement and enrich the details of our wavelet- and climate-based analyses, while such analyses above informed the form of our Bayesian model. During the model development period (2010–2017), we estimated strong model generalisability due to accurate estimated out-of-sample trends. The leave-one-time-point-out posterior predictive distributions' 95% credible intervals contained 82.3% of the observations (indicating a reasonably calibrated model) and the corresponding posterior median explained 66% of the variation in the DIR (i.e. $R^2 = 0.66$).

For model-based climatic inferences across the studied period (2010–2021), we analysed risk profiles (Figure SI C.21), where relative risk of greater DIR was measured with respect to the risk induced by the mean average of each climatic variable. Higher (lower) values of the two-month-rolling-averaged, monthly average of daily minimum temperature were associated with elevated (diminished) relative risk of human disease incidence. We also estimated a consistent risk profile for the cumulative drought indicator with elevated DIR risk for extreme values (large surpluses or deficits of accumulated precipitation). Intermediate values of the indicator were associated with lower relative risk. In terms of the two-month rolling average of monthly precipitation, we estimated greater risk of disease incidence for values between 150mm and 200mm at lags of approximately one month, whilst extreme large precipitation values, except at very small time lags, were associated with diminished risk of DIR. Finally, we estimated a complex risk profile for the sea surface temperature indicator (ICEN) as extreme large values (El Niño events) and extreme small values (La Niña events) were each associated with both elevated and diminished risk at different time lags across four months. It was difficult to infer the mechanisms behind some of the complex behaviour in the risk profile for the ICEN (see Discussion).

Ensemble forecasting

We deployed our Bayesian model in an expanding window, forecasting environment (excluding current and all future months' data from model training) for each month in the model development period (2010 to 2017), and found inferior

Table 4.3: **Different models' results for forecasting log(Cases + 1) in the testing period of 2018 to 2021.** We summarise proper scores, predictive performance, calibration, and outbreak detection capabilities of forecasting models. Models are arranged by mean weighted interval score (WIS), from best (lowest) to worst (highest). Further summary statistics for WIS are presented in Table SI C.14, and implied similar model rankings. Bias measures the bias of each models' predictive quantiles, PI coverage measures the coverage of the 95% prediction intervals (PIs), R^2 denotes the proportion of variation in log(Cases + 1) explained by the models' posterior median estimates, and RMSE is the corresponding root mean square error (RMSE) for log(Cases + 1) by using the posterior median. For detection of outbreak onset in each province in each year, *Sensitivity*, *Specificity*, and *AUC* represent the True Positive rate, True Negative rate, and Area Under the Curve in terms of forecasting skill for correctly classifying the first possible outbreak of DIR $\geq$ 50 per 100,000. These metrics are visualised in Figure 4.11 a for model comparison purposes. Table SI C.9 is the analogous table which depicts outbreak detection for DIR first surpassing a threshold of 150 per 100,000, while Table SI C.12 presents the models scored on a DIR level. Details of the model abbreviations are provided in Table 4.2 (where similar model rankings were obtained). All ensemble models are highlighted with an asterisk.

	Model	WIS	Bias	PI Coverage	R^2	RMSE	Sensitivity	Specificity	AUC
1	Median *	0.34	0.03	95.2%	0.74	0.81	81.8%	94.3%	0.88
2	Median-NoBase *	0.35	0.02	94.8%	0.73	0.82	81.8%	95.7%	0.89
3	Median-NoBayes *	0.35	0.06	94.8%	0.74	0.82	81.8%	87.2%	0.85
4	EW-Mean *	0.35	0.17	87.8%	0.75	0.79	72.7%	95.9%	0.84
5	Median-NoCov *	0.36	0.04	95.2%	0.73	0.83	81.8%	91.0%	0.86
6	Ew-Mean-NoBayes *	0.36	0.21	87.5%	0.74	0.81	63.6%	97.3%	0.80
7	EW-Mean-NoCov *	0.36	0.14	91.2%	0.74	0.81	81.8%	86.5%	0.84
8	Trained *	0.36	0.07	92.4%	0.72	0.84	81.8%	95.4%	0.89
9	EW-Mean-NoBase *	0.36	0.16	87.1%	0.74	0.80	81.8%	94.7%	0.88
10	SARIMA	0.36	0.03	93.5%	0.73	0.82	63.6%	96.1%	0.80
11	Prov-Trained *	0.37	0.05	91.1%	0.72	0.85	72.7%	89.9%	0.81
12	Baseline	0.38	0.04	97.3%	0.74	0.83	90.9%	45.0%	0.68
13	TCN	0.39	0.11	91.2%	0.70	0.87	72.7%	96.6%	0.85
14	TimeGPT	0.41	0.08	88.1%	0.69	0.91	90.9%	91.7%	0.91
15	TimeGPT-NoCov	0.42	0.12	85.6%	0.67	0.92	72.7%	91.5%	0.82
16	Bayes-Climate	0.52	-0.20	87.6%	0.58	1.09	54.5%	90.1%	0.72

results versus the leave-one-time-point-out analysis. Irrespective of assessment on a logarithmic or DIR level (see Materials and methods), we identified model calibration issues due to frequently underpredicting posterior predictions and hence, poor interval and quantile coverage values (Figures SI C.28 and SI C.29). This suggested the Bayesian climate-based model alone would likely be inappropriate and unreliable for probabilistic forecasting, and supported our motivation for ensemble frameworks. While it would be unwise (due to bias from model tweaking after seeing the data) to focus heavily on results from the model development period, we found stronger performance for each of our other ensemble candidate members (see Section 4.2.2), including assessments of the whole predictive distribution (the WIS), pointwise accuracy (e.g., R^2), PI and quantile coverage, and bias (Table SI C.8). For example, the simple yet flexible SARIMA models improved the calibration (e.g., quantile and interval coverage in Figures SI C.28 and SI C.29), the accuracy of pointwise predictions (e.g., $R^2 = 0.79$, see Table SI C.8 and Figure SI C.23), and the overall distributional accuracy (WIS). Our deep learning models, particularly TimeGPT (the foundational time series model) which incorporated climatic covariates, also yielded strong forecasting performance relative to the baseline forecaster, the models without climatic covariates, and the Bayesian climate-based model (Figures SI C.22-SI C.27, Table SI C.8). Simple, untrained ensembles also consistently produced accurate pointwise estimates and strong calibration and sharpness.

Our four-year testing window of 2018 to 2021 was an environment to assess performance in a setting reflective of a real-world application, as all models (including hyperparameters, covariates, and ensemble weights) were pre-defined from only historical data. We found strong predictive performance in forecasting cases for each of our ensemble models (Tables

4.3, SI C.9), as captured by WIS, PI and quantile coverage, and pointwise accuracy metrics (R^2 and RMSE – Root Mean Square Error). We found similar strong performance of ensemble models for forecasting the annual first onset of DIR reaching thresholds of 50 (or 150) per 100,000 (Figure 4.10 a), and more generally for forecast classification of months with DIR above/below these thresholds (Figure 4.10 b). The individual models handle different types of uncertainty in different ways (e.g., uncertainty in data generating processes, random error terms, parameter estimates etc.), yet individually do not account for all possible sources of uncertainty and thus, the coverage of their 95% PIs is less than the claimed coverage (95%). In contrast, relative to other ensembles and individual models, ensembles based on the median of individual models' quantiles produced statistically valid PIs and consistently strong forecasting ability (Figure 4.10). This was captured by consistently sharp and well-calibrated, albeit relatively conservative, PIs (Figure SI C.41), low bias, and better WIS values (e.g., pairwise comparisons of WIS values in Figure SI C.37 and Table 4.3). We observed these trends across provinces (Figure 4.11 e) and over time (Figure 4.10 c), and for different magnitudes of epidemics (Figure SI C.39 and Table SI C.15). The superiority of forecasts from ensemble models, particularly median-based ensembles, was also observed when we ranked models on their forecasts of DIRs. Across assessments on logarithmic and DIR scales, median-based ensembles consistently dominated the model rankings based on score ratios, outbreak detection, and pointwise accuracy metrics (Tables 4.3, SI C.9, SI C.12, SI C.13 and Figures 4.10 c, 4.11 a and SI C.37).

While the assessments of WIS values (and other statistical measures) provide statistically rigorous measures of the whole predictive distribution, we also focus on the insights of the forecasts for public health authorities. For instance, the epidemiologically naive baseline model's predictive distributions could be relatively close in distribution to the observed cases (if true distributions do not change drastically in short periods), yet this model fails to offer valuable information about oncoming outbreaks (as it forecasts the past observed cases). So, looking at detection of the annual first onset of DIR reaching thresholds of 50 (or 150) per 100,000 (Figures 4.10 b, 4.11 a and 4.11 c), the median-based ensemble models produced AUC values of between 0.85 and 0.89, which are superior to i) the naive baseline forecasting model (AUC: 0.68) and ii) other individual models. For example, if authorities used the median-based ensemble of all individual models, 81.8% of first onsets (of reaching DIR of 50 per 100,000) would have been correctly detected one month ahead of time with a corresponding false alarm rate of 5.7%. We found similarly strong performance of ensemble models for classifying whether DIR in a province in any given month would be at or above the thresholds of 50 (or 150) per 100,000. Here, nearly all forecasting models demonstrated strong classification skill (e.g., AUC of between 0.79 for DIR $\geq$ 50 per 100,000, see Figure 4.11b and Table SI C.10), likely in part due to the easier nature of predicting the next month being at/above a threshold given that we know the most recently observed high DIR level.

In terms of further relative comparisons, forecasts of ensemble models which included climatic covariates outperformed the analogous models without covariates (denoted NoCov in e.g., Table 4.3), while trained ensembles were outperformed by untrained median-based ensembles (see Discussion) across both logarithmic and DIR scales (Figure 4.11 a). Ensemble models also outperformed almost all individual forecasting models (e.g., 95% PI coverage and WIS), and model rankings were similar across DIR and logarithmic scales (Tables 4.3, SI C.9).

4.2.4 Discussion

We have introduced a new multi-model framework (Figure 4.7) which enables multi-model evidence synthesis; as a researcher can explain the dynamics and drivers of past epidemics and robustly estimate likely future short-term epidemic trajectories.

Wavelet analysis and climate-based modelling

Using wavelet methods allows us to identify epidemic dynamics and drivers that vary in different temporal resolutions, times of the year, and/or in different geographies. We found strong similarities in both the seasonality and epidemic curves

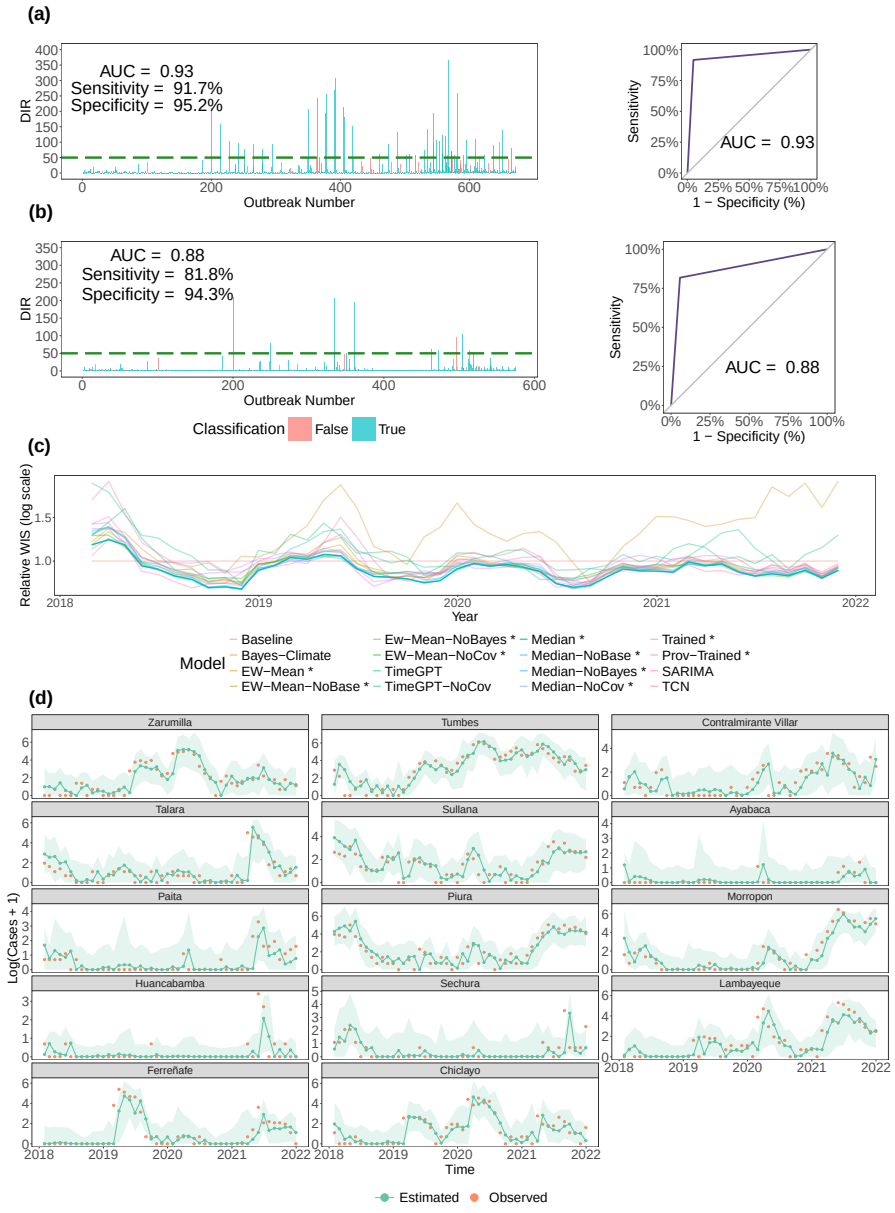


Figure 4.10: **Performance of median ensemble forecasting model:** This ensemble forecasting models' probabilistic forecasts were generated by taking the median of all six ensemble forecasting models' prediction at the 23 quantile levels (see Materials and methods). (a): Visualisations of forecasting ability for classifying whether the coming month will experience a DIR of $\geq$ 50 per 100,000. AUC denotes the area under the receiver operating characteristic (ROC) curve, and higher values indicate greater classification skill. (b): Analogous visualisation for classifying whether the coming month will be the first month of the year in a province with DIR of $\geq$ 50 per 100,000. (c): Relative Weighted Interval Score (WIS for log(Cases + 1)) is shown for all forecasting models in centred, rolling windows of three months, with the median model's values highlighted. Relative WIS is an individual forecasting models' WIS divided by the baseline forecasting model's WIS, where values below one indicate greater forecasting skill relative to the baseline forecaster. The corresponding summary statistics for log(Cases + 1) over time are visualised in Figure SI C.39. (d): For each province (sorted in descending order by latitude from north to south), we visualise the median model's predictions (for log(Cases + 1)) where green points denote the posterior median and orange points denote the observed values. The green shaded intervals denote the 95% prediction intervals (PIs). 95.2% of all observations lay within these 95% PIs. The posterior median explained 74% of the variation in log(Cases + 1) i.e. $R^2 = 0.74$. Y-axes are province-dependent due to the varying sizes of the dengue epidemics observed.

across the provinces, and similarities were pronounced during epidemic years with major outbreaks. It is unsurprising that northerly provinces, which experienced higher DIRs, possessed more synchronous and more coherent (annual and

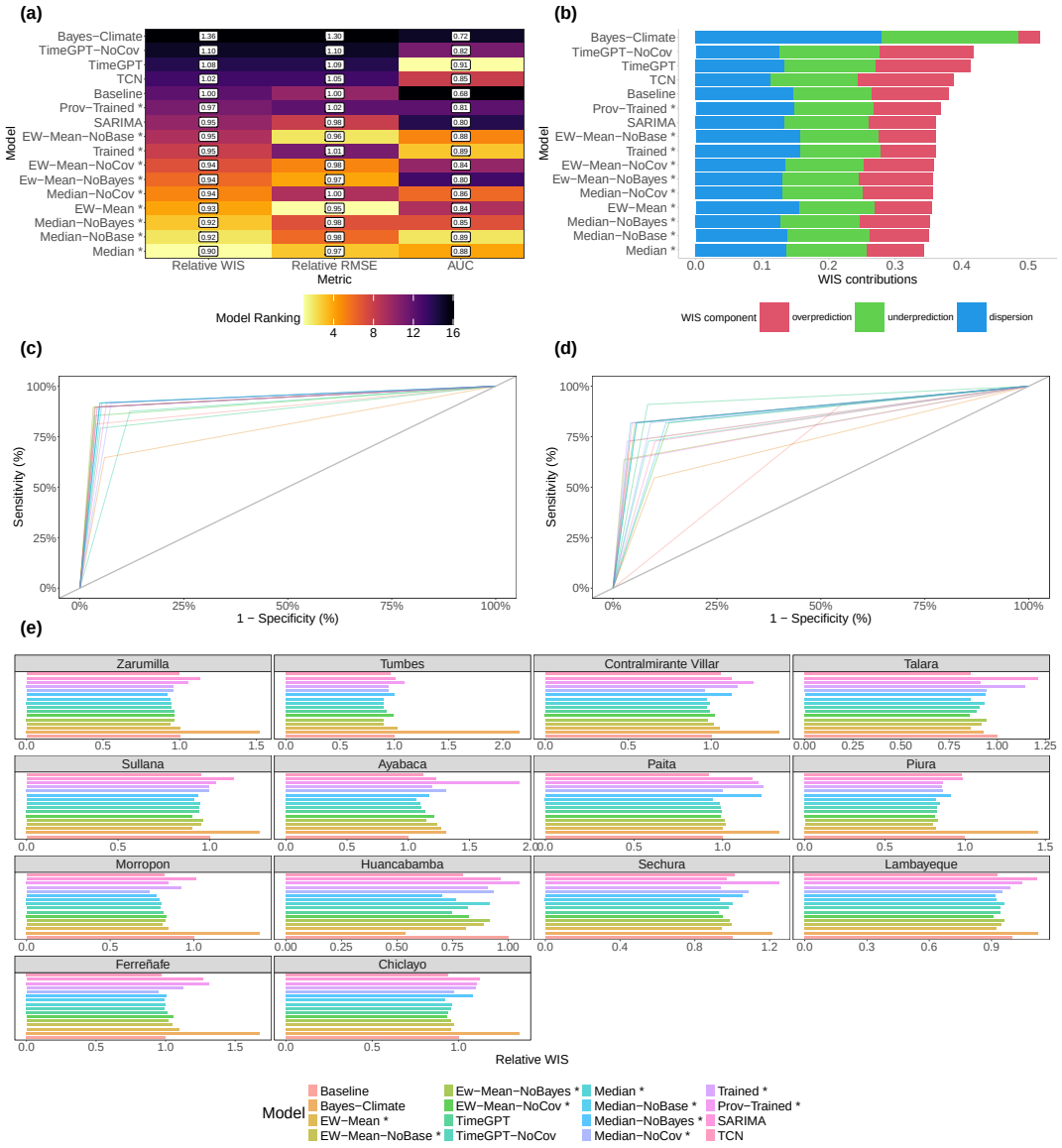


Figure 4.11: **Comparing relative performance of forecasting models in the testing period of 2018 to 2021.** We compare performance of our 16 model frameworks (see Table 4.2 for abbreviations) in terms of their predictive performance for forecasting log(Cases + 1). (a) Model rankings by different metrics, where relative root mean squared error (RMSE) and mean relative Weighted Interval Score (WIS) are defined relative to the baseline model. Lower values indicate greater predictive performance for these relative metrics, but higher AUC values indicate greater forecasting skill for correctly classifying the first months with DIR $\geq$ 50 per 100,000 one month ahead of time. (b): WIS is decomposed into overprediction, underprediction, and dispersion for each of the forecasting models. (c): Receiver operating characteristic (ROC) curves by forecasting model for correctly classifying months with DIR $\geq$ 50 per 100,000 one month ahead of time. (d): Analogous ROC plot to (c) for detection of first month with DIR $\geq$ 50 per 100,000 one month ahead of time. (e): Relative WIS is shown for all forecasting models per province where provinces are sorted in descending order by latitude from north to south. Relative WIS is an individual forecasting models' WIS divided by the baseline forecaster's WIS for that province. X-axes are province-dependent due to the varying sizes of the dengue epidemics observed.

multiannual) dengue incidence cycles. Also, the strongest wavelet coherence between climatic conditions and dengue cycles occurred in more westerly and northerly provinces (all of which were coastal provinces). Consistent with wavelet analyses in Southeast Asia[539], such significant peaks in wavelet coherence and cross-wavelet power were estimated during years with larger epidemics, and were widespread, all of which is indicative of an amplifying role (greater epidemic similarity) played by climatic forcing in large epidemic years. The spatially shifting impact of the ICEN during such

years (from west to east and north to south) may reflect initial climatic risks imposed on coastal provinces when an El Niño (or La Niña) event initially occurs, dengue incidence climbs, and subsequent climatic conditions and human mobility induce transition of outbreaks further inland and further south. From a policy perspective, in future work, the roles played by both climatic forcing and spatially shifting influences could be combined with seasonal climate forecasts (e.g., GloSea5[567] was used for dengue forecasting in Vietnam[162]) to enhance early warning detection. For the causes of epidemic synchrony between provinces, we identified large outbreak years, greater pairwise products of populations and shorter pairwise distances as having enhancing effects. This may reflect greater mobility and greater connectivity driving enhanced epidemic synchrony. Spatially, shorter pairwise distances had more significant effects on both synchrony and coherence in more northerly provinces, which may be indicative of greater connectivity (although this may be confounded by similarity in other environmental conditions). The more uncertain estimated effects of climatic conditions on synchrony may be due to our limited spatial focus (only fourteen provinces) and/or potential confounding in our models of synchrony.

The estimated elevated risk for anomalous precipitation deficits and accumulations is aligned with the well-documented impacts of water shortages and heavy rainfall[78,499,500]. Drought-induced precipitation shortages, particularly in areas with poor water supply, can induce greater risk of dengue incidence due to changes to water storage practices (including more containers around the home), creation of potential egg-laying sites (for female *Ae. aegypti* who prefer artificial water storage containers), greater survival capacity of *Ae. aegypti* eggs in dry conditions, and/or more concentrated breeding in the few remaining suitable habitats[84,526,527]. Similar elevated drought-induced risk has been identified in both Vietnam and Brazil[78,79]. Meanwhile, the effects of heavy rainfall are mainly due to large-scale accumulation of stagnant water, which yields creation of abundant breeding sites, and increases to vector populations. The risk of disease incidence was estimated to peak after a one-month lag for the two-month average of monthly precipitation, whilst we estimated diminished risk for the heaviest precipitation (above a two-month average of 200m). Diminished risk has been uncovered in Asian nations including Singapore, Malaysia, Sri Lanka, and Thailand[353,568]. Our final climatic variable (the ICEN) was estimated to have a complex risk profile. Whilst the estimated diminished risk for extreme indicator values at short time lags make intuitive sense (due to time taken for onshore climatic conditions of provinces to eventually change), protection (from risk) for greater absolute values at longer time lags seem counter-intuitive. The latter issue may due to the limited size of our data and/or using our predictive model to disentangle real-world climatic inferences (which may be subject to potential confounding). The issue of potential confounding (between variables) is applicable to each climatic variable, whilst external, unobserved confounders may also have affected our conclusions and thus, we indirectly estimated such unobserved confounding effects (via unstructured spatial effects and yearly random effects).

There are several limitations to both our retrospective and forecasting analyses. These include concerns about the quality, volume, and type of data available, ranging from a lack of information on reporting rates to a lack of serotype data (See Appendix C.8 for further discussion). We attempt to compensate for potentially changing serotypes and reporting rates with time-varying assessments of climatic-incidence relationships in our wavelet analyses, year-to-year random effects in our Bayesian model, and consistent updating of all hyperparameters in each forecasting model (including ensembles). Nevertheless, consistent with the WHO's cited lack of data on circulating serotypes[569], we advocate for future collation of age-structured dengue cases and viral sequencing data, alongside future seroprevalence surveys across at-risk areas, all of which would provide insights about immunology and virology issues (such as multitypic immunity, genomic diversity, and repeat infections) across space and time. Similarly, while often logistically and economically challenging to collect, including mosquito abundance data (conditional on appropriate alignment of likely different spatial scales[190,218]) would ideally allow us to build a more mechanistic model and enhance exploratory power. Data regarding the role of the human are sometimes overlooked yet vital for vector-borne transmission, and given the high spatial synchrony between provinces, we also advocate for enhanced availability of human mobility data over time. Understanding how continuous water supply modulates climatic-incidence relationship is another important investigation that would be enabled by further data.

Raising the necessary caveat that more data would not guarantee enhanced accuracy/representativeness[249], quantifying the additional value of these data streams (e.g., using Value of Information analyses) for different policy questions (Figure 4.7) and different settings will be important for encouraging the potentially costly data collection in resource-limited settings.

Forecasting analysis

Moving from past to future dynamics, in a four-year testing window, we demonstrate a new probabilistic ensemble framework for forecasting monthly dengue cases up to one month ahead of time.

The reliable forecasts from ensemble frameworks demonstrate benefits of combining the strengths of multiple models from different disciplines. Each of these models are "wrong" in different ways, and yet each can handle different aspects of forecasting dengue cases. The foundational time series model (TimeGPT) included in the current work leverages strengths from a transformer architecture, conformal inference, and training on over 100 billion data time points. The TCN model uses representation learning and dilated causal layers in its deep neural network architectures. In doing so, these deep learning models can explicitly account for temporal structure (from seasonality to heteroscedasticity) and handle nonlinearities in the data. While deep learning models can be computationally expensive, here, there are little to no additional (e.g., versus Bayesian hierarchical models) computational barriers for operationalising these models in potentially resource-limited settings as all our computations were performed locally on a laptop (with each provinces' monthly forecasts produced in under one hour, irrespective of model). However, the fact that deep learning models do not always outperform simple-yet-flexible and robust SARIMA models is interesting and perhaps unsurprising. Dengue epidemics are somewhat predictable in terms of their highly seasonal patterns (e.g., dengue cycles based on climatic conditions) and their very short-term trajectories with highly autoregressive patterns (e.g., we expect the next month's DIR to somewhat similar to the current month's DIR). So, unlike some previous studies which either benchmark against models only within the deep learning field or do not allow the SARIMA model's hyperparameters to be updated as new information becomes available, our rigorous benchmarking reveals similar performance of these statistical time series models with deep learning methods (which may have increased propensity to overfit due to the far larger numbers of model parameters). The similarity in performance is implied by comparable (or superior) values for PI coverage, bias, WIS, R^2, and outbreak detection metrics. Similar results have been observed for forecasting of COVID-19 cases and deaths[223] and in other unrelated forecasting tasks with strong, seasonal patterns (such as retail demand and wholesale food prices)[570,571]. Of course, this observation comes with the caveat that future work should assess how this predictive performance varies at longer forecast horizons and finer spatial and temporal resolutions (e.g., weekly), all of which are of great importance to public health authorities. We think that our extensive model and ensemble development represents a stepping stone towards such additional objectives. Future research could consider methods for achieving multi-step forecasts (e.g., recursive posterior predictive sampling, direct multi-step forecasts, and sequence modelling for Bayesian, statistical time series, and deep learning transformer-based models respectively) and/or incorporating new and increasingly accurate short and medium-term weather forecast data (which often use artificial intelligence, e.g., diffusion models or graph neural networks) to expand dengue case forecast horizons[53,285,286].

Importantly, however, our deep learning models do always outperform the Bayesian climate-based model and improve our ensembles' forecasts when included alongside other individual models. The latter trend is reflective of a more general observation of the outperformance of ensemble models, compared to individual models (similar to previous forecasting studies for COVID-19, influenza, and dengue[169,433,572]). Intuitively, similar to our retrospective analysis, it is unreasonable to expect a single approach (wavelets and Bayesian models also have differing strengths and weaknesses) to capture all aspects of current or future epidemic dynamics. This is often due to data distribution shifts, and learned patterns may therefore vary substantially across space and/or evolve over time. These variations may be why our province-specific deep learning forecasting models (TCN and TimeGPT) with climatic covariates and statistical time series model (SARIMA)

forecasted more reliably than our semi-mechanistic Bayesian climate-based model which estimated parametric (climate-to-incidence) relationships for all provinces and time points jointly. These trends are similar to previous poorer short-term predictive performance of models with mechanistic structure (and stronger performance of ensembles) in analyses of forecasting new dengue and COVID-19 cases[169,223], and may capture the difficulties in using mechanistic models for forecasting imperfect (noisy, underreported, and delayed) case data. Future work may consider pairing our framework with nowcasting approaches which can infer realised events which have not been reported yet (see[114,366,469,573] for recent methods and applications). Despite periods of inferior performance relative to the baseline model (Figure SI C.39), the forecasting ensembles (e.g., median-based ensemble) generally provided more stable performance over time and across space (Figure 4.11), both during periods of low and high dengue incidence. We deduce that a single model alone is likely insufficient for answering all of our research questions, yet we can answer different research questions robustly by using and where appropriate, unifying, several models. Our aim here was not to include every modelling approach in a single framework, but rather to identify which models/approaches were appropriate for different tasks (e.g., probabilistically explaining past epidemic drivers vs predicting future trajectories).

Our forecasting results are not an upper bound on what is possible, as i) we have retained a preference for realism in our independent testing window (and we could also incorporate future covariates), ii) our pipeline can readily incorporate many more exploratory and forecasting models, and iii) future work could investigate further feature selection, optimal training windows for individual models and for weighting of models, as well as model weights per individual quantile. These future works may help to improve the performance of the ensemble in provinces which experience rare, sporadic dengue cases (e.g., Ayabaca). Our focus here has instead been to showcase that we can, for any dengue epidemic, blend, rather than overemphasising, the benefits of forecasting from single modelling domains.

Our work also highlights the importance of not focusing on a single, potentially misleading performance metric. We must consider what information the forecasting metric provides for public health authorities. While used for forecast evaluation in many dengue studies to date (e.g.,[555,556]), pointwise metrics obscure the pervasive uncertainty when we predict the unknown future of a dengue epidemic. We could, for example, obtain a high R^2 by just predicting the most recent value, while providing little useful information for public health authorities. Similar to other recent forecasting studies (e.g.,[58,162,172,433]), we therefore propose proper scoring rules which measure distributional accuracy (WIS) and account for deficiencies of pointwise metrics by simultaneously evaluating calibration and sharpness. This is why we focus on WIS, yet not exclusively, as i) we could again obtain reasonably good WIS values by simply forecasting past value(s) (as drastic predictive distributional shifts may be rare in short time horizons) without conveying much information, and ii) we aim for our outputs to be easily interpreted and used by public health authorities. For example, for quick space- and time-specific forecasting model comparisons, we signpost Figures 4.10 d and 4.11 e as visually helpful guides. All of this is an important contribution of our work, as we complement proper scoring rules (which measure the whole distribution) with both relative WIS values (to reflect forecasting skill above the baseline) and measures of how well our forecasting models can detect specific aspects of epidemics, such as season onset. The WIS (and/or relative WIS) may be a statistically principled measure for our evaluations, yet when communicating our outputs, such a metric may matter so little to so many. In particular, the WIS approximates the CRPS – the simplest single measure of the overall value of a forecasting model for the entire uniform family of potential users. Ideally, metrics should also be interpretable, actionable, and public-health-focused (e.g., user- and/or event-focused), and so our outbreak threshold metrics refocus evaluation on areas of the predictive distribution that are vital for decision-making. These new outbreak thresholds (and the associated optimisation procedures) are simple yet generalisable across geographies of different sizes and across different spatial resolutions (e.g., coarser or finer resolution analyses).

In terms of further public health implications, it is interesting that we can still often obtain reliable probabilistic forecasts without climatic covariates. This may be related to the one-month forecast horizon as given that we know the current

level of the epidemic, the next month's cases may not require the climatic information for reliable forecasts. On the other hand, when making longer-term seasonal forecasts and inferring epidemic drivers as in our retrospective analyses, climatic covariates are crucial (e.g.,[162]). Another benefit of our approach is the computationally cheap nature of our forecasting pipeline. Working with 23 predictive quantiles per observation to summarise the predictive distribution saves a modeller from processing potentially thousands of predictive samples. Finally, ensembles with untrained weighting outperform those which use trained weighting, likely in part due to weight uncertainty and/or the restricted spatial and temporal focus of our study. More recent trained weighting approaches should be investigated in future work[572].

The next step will be operationalising this multi-model approach (Figure 4.7) into a decision support tool for the Peru CDC – a computational tool that iteratively downloads data and produces retrospective analyses, probabilistic forecasts, and associated metrics. We emphasise that the co-development of this tool with the Peru CDC does not end after detailed retrospective and forecasting analyses. Subsequent steps for operational success will include synthesising evidence, determining the necessary details for communicating modelling results, and regular dialogue for updating models and co-defining new policy questions. Maintaining such collaboration and clarity of audience will be crucial for satisfying the 3-U (useful, usable, and used) requirements of a practical, successful digital tool[280]. This may be aided by adopting lessons from policy-modelling experiences in previous infectious disease settings (e.g., COVID-19 and mpox[18,428]). In doing so, we believe that our framework and recommendations (possibly paired with decision theory, see[160]) can satisfy the necessary usefulness principle for a digital prediction tool[280], yet future work (e.g., surveys and involvement of social scientists) could explore i) usability; whether our custom outputs are interpretable (for different users involved at different stages of the decision-making process), and ii) usage; evaluation of effectiveness.

Conclusions

The key advance of our research is to introduce a new framework and multi-model way of evidence synthesis; robustly analysing dengue epidemic dynamics across space and over time. We introduce several state-of-the-art methods across different disciplines (in stand-alone and unified frameworks), yet their usage alone is not our objective, as we complement these methods with traditional, reliable modelling techniques to produce a computationally cheap and generalisable multi-model approach that emphasises public health utility throughout. The outputs of our retrospective and forecasting techniques are designed to be interpretable, robust, and comprehensive, and can be used to inform long-term public health planning and real-time decision-making for dengue epidemics.

Data Availability Statement

The dengue incidence surveillance data are publicly available from the National Centre for Epidemiology, Disease Prevention and Control (Peru CDC) in Peru's Ministry of Health. We sourced these from `https://www.dge.gob.pe/salasituacional`. Mid-year population estimates are made available by the National Institute of Statistics and Information of Peru at `https://www.inei.gob.pe/media/MenuRecursivo/indices_tematicos/proy_04.xls`. The WorldClim monthly historical climate data are available at `https://www.worldclim.org/`. SPI-6 data from the European Drought Observatory are available at `https://jeodpp.jrc.ec.europa.eu/`. The El Niño indices of the ONI and ICEN are available respectively from the NOAA (`https://origin.cpc.ncep.noaa.gov/`) and the Geophysical Institute of Peru (`http://met.igp.gob.pe`). All code and data used in our analysis are available at `https://github.com/cathalmills/peru_dengue_province/`.

Chapter 5

The time- and space-varying roles of human mobility in shaping urban dengue outbreaks

Cathal Mills, Geziel dos Santos de Sousa, Antonio Silva Lima Neto, Vasco Furtado, Sen Pei, Moritz U. G. Kraemer, and Christl A. Donnelly. The time- and space-varying roles of human mobility in shaping urban dengue epidemics, June 2025. URL https://www.medrxiv.org/content/10.1101/2025.06.03.25328731v1. Pages: 2025.06.03.25328731

This chapter presents a manuscript, currently under review at *PNAS Nexus*[47], that introduces a multi-model evidence synthesis approach to quantify the roles of human movement in urban dengue outbreaks across space and time.

Abstract

Dengue outbreaks continue to cause large morbidity and mortality globally. Previous work has shown that human movement alongside weather, climatic, ecological, and socioeconomic factors influence outbreak size, persistence, and geographical spread. However, the relative importance of human movement is unclear for the establishment, rapid expansion, and persistence of dengue in urban settings. Relatedly, the extent to which neighbourhoods differentially influence outbreaks over time is unclear. To address these gaps, we developed a multi-model framework that integrates Bayesian hierarchical modelling and data-driven, deep-learning-based approaches to parameter inference and out-of-sample probabilistic predictions of dengue in Fortaleza, Brazil. We apply our framework to dengue outbreaks at the neighbourhood level, using epidemiological surveillance data alongside public transportation data. We use information criteria, scoring rules, and explainability metrics to measure the models' performance and ultimately explain how human mobility shapes disease dynamics. We find that human mobility is a consistent driver of dengue outbreaks over the five-year period studied. We also find that human movement is a more important driver of transmission dynamics between neighbourhoods than transmission within them, both within and across dengue seasons, and that spatially, the relative importance of different neighbourhoods to transmission elsewhere is relatively constant over time. Our framework highlights the time- and space-varying roles of human mobility and is applicable to outbreaks of other infectious diseases and to other questions of relative epidemic drivers.

Author contributions

- Conceptualization: **C.M**

- Methodology: **C.M**; M.U.G.K; C.A.D, S.P
- Data curation: **C.M**; G.S.S
- Data visualisation: **C.M**.
- Writing original draft: **C.M**; C.A.D.
- Writing final draft: **C.M**; A.S.L.N; V.F; S.P M.U.G.K, C.A.D.

5.1 Introduction

Dengue is a vector-borne disease caused by the dengue virus (DENV) that has four distinct serotypes and is spread by *Aedes* mosquitoes, transmitted between humans primarily by *Ae. aegypti* and to a lesser extent, *Ae. albopictus*. Dengue is the most rapidly expanding mosquito-borne disease, approximately half of the world's population is currently at risk of infection [31,67], and the incidence and distribution are expanding [38]. Faced with the growing public health challenges associated with severe dengue, there has been an increasing effort to improve our understanding of the biological, human, and environmental influences on dengue epidemics around the world. Modelling is one way to provide such an understanding, and a wide range of approaches have been used, including statistical, mechanistic (or mathematical), machine learning (ML), and deep learning frameworks. Given the climatic sensitivities across both the pathogen (DENV) and mosquito vector [80,87,88], a common research objective has been to quantify how transmission dynamics have been shaped by recent climatic conditions and their interactions with human influences (e.g., urbanisation, human mobility, and socioeconomic factors [78,79,195]) and epidemiological influences (e.g., serotype replacement, serotype competition, and population immunity [438,574]).

In recent years, human movement has been identified as an important driver in dengue epidemic dynamics. In Pakistan, a mechanistic modelling study used mobile phone data and climate data to reveal that the inclusion of human mobility improved predictions of the spread and timing of dengue epidemics [85]. Mechanistic approaches have similarly been used in southern China and Singapore to model dengue spread within and across cities [235,575,576,577]. Statistical frameworks, such as Bayesian hierarchical models, have generally incorporated the effects of human mobility using either human movement matrices, distance-based approximations, or neighbourhood-based approximations [79,200]. In the city of Fortaleza in Brazil, both mobility-informed artificial neural networks and mechanistic models were employed at a fine spatial resolution (neighbourhood level). The inclusion of human mobility (bus transportation) data again yielded improvements in prediction accuracy [86]. A separate predictive modelling study [233] used data from the COVID-19 pandemic to identify human mobility as a crucial factor in shaping dengue epidemic dynamics in Fortaleza. At a longer-term horizon, results from Harish et al. [227], based on a hierarchical model of geographical expansion, indicated important synergistic roles for short- and long-distance human movement, alongside suitable environmental conditions, for the geographical expansion of dengue in Brazil and Mexico.

However, while the increasing availability of mobility data in different guises has enabled such modelling studies approaches (which extend beyond mobility approximations [238]), we still have a poor understanding of how the relative importance of human movement in dengue outbreaks varies, within and across seasons and across space. Such time-varying effects may be important, as genomic studies indicate that once dengue is established, reimportations remain important [190]. We consider here when, where, and how human movement shapes urban dengue outbreaks at the neighbourhood level in a city (Fortaleza, Brazil) where dengue is endemic for over 30 years. Research questions we ask are: In an urban context, how does the role of human mobility change through time? How does human mobility shape the epidemic dynamics once the outbreak is under way? How does the relative importance of human mobility change across years? How do neighbourhoods differentially impact the city-wide dynamics of dengue over time? We investigate these questions using

mobility-informed statistical and deep learning models, alongside statistically principled multi-model development and new evaluation procedures. Our methodology for mobility-informed modelling allows us to infer the high-resolution time- and space-varying relative importance of human mobility for driving urban dengue outbreak dynamics, and quantify the impacts of human mobility on source-sink dynamics of neighbourhoods. Looking ahead, our framework can provide the insights necessary for public health authorities to design more targeted and optimally timed public health interventions, both for dengue and other infectious diseases.

5.2 Materials and methods

5.2.1 Study area and data

Fortaleza is a coastal city in north-east Brazil with a population of approximately 2.4 million people[578]. It is the fifth-largest city in Brazil and consists of 119 neighbourhoods/bairros (Figure 5.1 A).

Dengue is endemic in Fortaleza. The first recorded dengue outbreak occurred in 1986 and was caused by the introduction of DENV-1. DENV-2 was first introduced into Fortaleza in 1992 , with the first major epidemic occurring in 1994 when approximately 660,000 dengue cases were reported in Fortaleza (approximately 84% of all cases in Brazil) driven in part by the suspension of mosquito control activities in late 1993 due to a cholera outbreak[579,580]. By 2012, all four serotypes of DENV had been detected in the city. Female *Ae. aegypti* are the primary vector of DENV transmission across Brazil. Suitable conditions, for the development of their larval habitat and DENV transmission, are provided by the highly urbanised Fortaleza neighbourhoods and the equatorial savannah climate (with distinct tropical rainy and dry seasons)[70,78,576].

To measure human mobility, we used previously published data for the public bus transportation system in Fortaleza; the most widely used mode of public transport with an estimated 700,00 people using the system every day[581,582,583]. The mobility data were available on a weekly basis for 2015. Using the coordinates of the origin and destination location, we obtained a directed measure of the number of flows ($f_w^{(i,j)}$) from neighbourhood i to neighbourhood j for week w. In the absence of further available data, we assumed the weekly mobility matrix (i.e the directed adjacency matrix) from 2015 applied to corresponding weeks in all other years (2007–2014). The mobility data were described extensively in previous work[86,581,582,583].

We sourced data on newly reported dengue cases from[86], which used reported cases from the Sistema de Monitoramento Diário de Agravos—SIMDA (Daily Disease Occurrence Monitoring System). These cases arise from diagnoses in hospitals or primary health care units. Note that dengue cases in Brazil can be confirmed solely on the basis of clinical and epidemiological criteria, without laboratory tests. The data were reported on a weekly basis (using diagnosis dates) for each of the 119 neighbourhoods according to the cases' home addresses.

For meteorological variables, we calculated daily mean temperature (at a 2 metre level) and daily total precipitation for Fortaleza using data from the ERA5-Land reanalysis dataset[282]. Due to data resolution, the same climatic conditions were used for each neighbourhood, and are generally characterised by year-round stable, suitable temperatures for dengue transmission and distinct rainy and dry seasons (Figure SI D.1).

5.2.2 Model development and fitting periods

To guard against optimistic bias and overfitting, we used data from 2007 to 2010 inclusive as a model development period. During this period, we selected combinations of covariates, developed model structures, and made other model choices based on information criteria (see below).

We then focused on data from 2011 to 2015 inclusive to test model performance and draw insights. Rolling windows

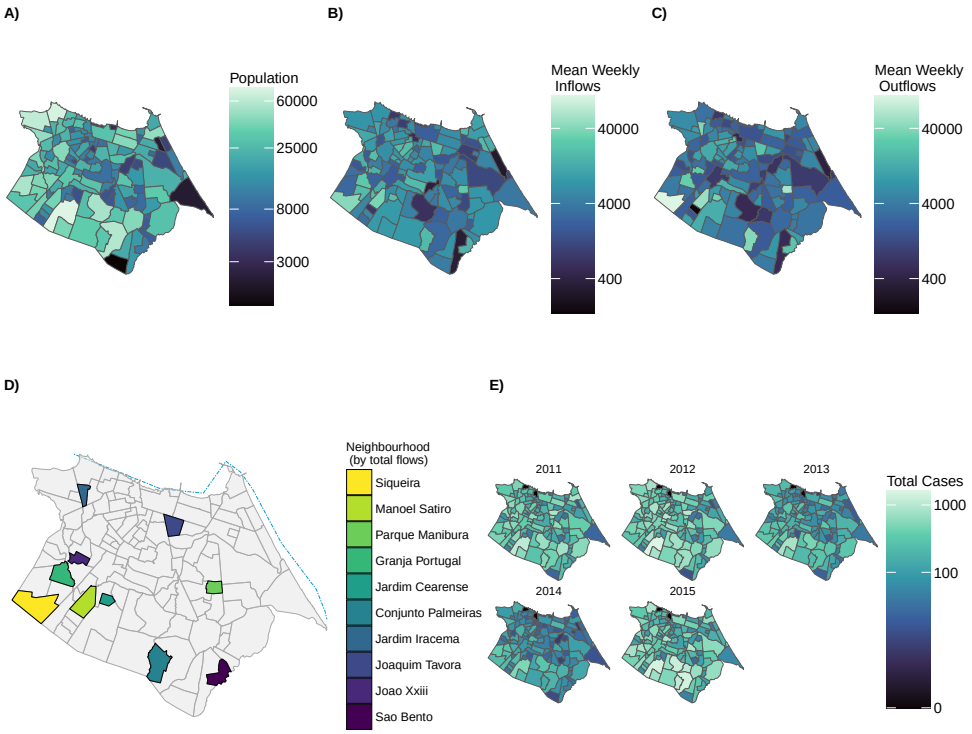


Figure 5.1: **Demography, human movement, and dengue cases in Fortaleza, Brazil:** Fortaleza is a coastal city in north-east Brazil, with a population of approximately 2.4 million people[578]. Within Fortaleza, there are 119 neighbourhoods (known as bairros), and all our data and analyses are at the neighbourhood level. **A)** The population (visualised on a logarithmic scale) of each of the 119 neighbourhoods, **B)** and **C)**: The mean weekly total number (visualised on a logarithmic scale) of inflows and outflows respectively for each of the 119 neighbourhoods. **D)** The neighbourhoods with the top 10 total number of inflows and outflows (visualised on a logarithmic scale) across all 52 weeks. Neighbourhoods are ordered in descending order from highest to lowest flows, and the blue line visualises the coastline of Fortaleza. **E)** The cumulative reported dengue cases (visualised on a logarithmic scale) in each neighbourhood per year of the testing period (2011–2015). The time series of cases within each year is visualised Figure 5.3 A–B, while a summary of cases per week are visualised in Figure SI D.3.

allowed us to estimate spatiotemporal variations in mobility effects without explicitly modelling the time variation for each observation, helping to guard against non-identifiability issues which could be caused by time-varying effects for each time point and neighbourhood.

After exploratory analyses (2007 to 2010), we chose a 17-week (approximately four-month) rolling window to allow sufficient learning of parametric relationships in our Bayesian hierarchical model, without losing excessive temporal resolution. The model learned about mobility influences within each rolling window, a procedure which also prevents averaging over potential data distribution shifts (or equivalently, changing mobility-incidence-rate relationships). For our deep learning models, we fitted these models using all information up to a given time t, due to the need for larger volumes of data to estimate covariate effects. Unless otherwise stated, all evaluation occurs in the rolling windows.

5.2.3 Bayesian hierarchical models

We developed Bayesian hierarchical approaches to modelling y_{it}, the reported neighbourhood-level (i) weekly (t) case counts of dengue. We used a Negative Binomial (NegBin) likelihood to allow for potential overdispersion in the case data.

We incorporated the graph structures from the weekly mobility data using a lagged mobility-weighted sum of all other

neighbourhoods' past dengue incidence rates (DIRs) per 100,000, which provides an estimated force of dengue importations. This approach was a computationally cheap (albeit limited, see Discussion) way to incorporate spatiotemporal structure that was informed by human mobility data. We used mobility flows and reported DIR from the previous week as inflow of infected individuals is less likely to result in new reported human cases within the reporting week (given the intermediate role of the mosquito vector, viral replication rates within the host and vector, and possible reporting delays).

We developed and fitted a range of candidate models (Appendix Section D.1) in R version 4.4.0 using the cmdstanr interface to Stan, a platform for Bayesian inference[458,584,585]. Using the R package loo, we estimated the out-of-sample generalisability of each model using ELPD-LOO; leave-one-out (LOO) estimates of the expected log pointwise predictive density (ELPD)[466]. Based on results for the model development period, we selected the following candidate model which assumes a fixed mobility parameter across space and a space-varying autoregressive parameter:

$$y_{it} \sim NegBin(\mu_{it}, \phi), \tag{5.2.1}$$

$$\log(\mu_{it}) = \log(P_{it}) + \log(\eta_{it}), \tag{5.2.2}$$

$$\log(\eta_{it}) = \alpha + \beta_{\text{Mobility}} \cdot \left(\sum_{j \in \mathcal{N}(i)} \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}} \cdot \log(\eta_{j,t-1}) \right) + \beta_{\text{AR[i]}} \cdot \log(\eta_{i,t-1}) + \beta_{\text{Year}[t]} + \beta_{\text{Nbr}[i]}, \tag{5.2.3}$$

$$\beta_{\text{Mobility}} \sim N(0, 1), \tag{5.2.4}$$

$$\beta_{\text{AR}} \sim N(0, 1), \tag{5.2.5}$$

$$\alpha \sim N(0, 1), \tag{5.2.6}$$

$$\beta_{\text{Year}} \sim N(0, 1), \tag{5.2.7}$$

$$\beta_{\text{Nbr}} \sim N(0, 1), \tag{5.2.8}$$

$$\phi^{-1} \sim \text{Half-Normal}(0, 1), \tag{5.2.9}$$

where μ_{it} denotes the mean parameter of the Negative Binomial distribution and ϕ is the reciprocal overdispersion parameter. We used a population offset $\log(P_{it})$ such that η_{it} represented the DIR per 100,000 people. We then modelled the (logarithm of) DIR with i) an intercept α, ii) an estimated force of dengue importations – a mobility effect controlled by fixed parameter β_{Mobility} and the lagged (logarithm of) DIR of all other neighbourhoods, $\mathcal{N}(i)$, weighted by the proportion of lagged inflows from such other neighbourhoods $\mathrm{f}_{t-1}^{(j,i)}$, iii) a potentially non-stationary and neighbourhood-specific autoregressive process of order one, controlled by $\beta_{\text{AR[i]}}$ (which allows for within-neighbourhood transmission caused by factors beyond mobility), iv) a year-to-year random effect (i.e., β_{Year} is a varying intercept for each year), and v) a neighbourhood-specific random effect (i.e., β_{Nbr} is a varying intercept for each neighbourhood). Modelling the effects on a DIR scale allowed us to disentangle each of these effects from population artefacts (e.g., large versus small neighbourhoods). Our goal was to understand the relative importance of human-mobility-driven transmission versus local transmission and the mobility-driven roles of different neighbourhoods in epidemic spread. We also explored explicitly accounting for climatic drivers (weekly total precipitation and average temperature), but the lack of neighbourhood-specific data meant that, causally, climatic covariates could not provide information for climate-driven local transmission dynamics (see Discussion).

A fixed parameter for mobility within each window does not assume that the effects of mobility (on each neighbourhood's DIR) are homogeneous across space and time, as i) both the human movement and human case data vary across space and time, and ii) the model can be fitted in rolling windows which allows spatiotemporal variations in the autoregressive term and the levels at which inflows of infected individuals generate new infections.

Mobility effects across space and over time

Using this model (eqs (5.2.1) – (5.2.9)), we inferred parameter-based estimates of i) how human mobility varies in terms of its relative importance within and across neighbourhoods over time, and ii) how human mobility impacts the time-varying, relative contributions (i.e., source-sink dynamics) that neighbourhoods make to the epidemics of other neighbourhoods.

We computed the following ratio, $r_i(t)$, for each neighbourhood i at time t:

$$r_i(t) := \frac{\beta_{Mobility} \cdot \sum_{j \in \mathcal{N}(i)} \left(w_t^{(j,i)} \cdot \log(\eta_{j,t-1}) \right)}{\beta_{\mathrm{AR}[i]} \cdot \log(\eta_{i,t-1})}, \tag{5.2.10}$$

where $w_t^{(j,i)} = \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}}$ is the weight assigned to neighbourhood i for the estimated inflows of infected travellers from neighbourhood j.

Eq. (5.2.10) measures the estimated time-varying relative importance of human mobility; contributions (to neighbourhood i's epidemic) from all other neighbourhoods, relative to neighbourhood's own autoregressive epidemic process, with values of $r_i(t)$ above (below) 1 indicating a greater (lower) relative importance of mobility-driven case importations. Using DIRs enabled comparison of the relative importance of human mobility across neighbourhoods of potentially different sizes. We compared how $r_i(t)$ differed across space (i) and over time (t), which helped us investigate the question of whether mobility's relative importance in driving dengue seasons is temporally homogeneous across space – i.e., does the importance follow similar patterns within and across seasons in different neighbourhoods?

We also estimated source-sink dynamics by deriving formulae for $\widetilde{C_j}$. This variable was standardised per window across neighbourhoods and reflects the relative contributions of a given neighbourhood j to the dengue epidemic dynamics in all other neighbourhoods. Values of $\widetilde{C_j}$ above (below) 0 indicate that a given neighbourhood has contributed more (less) to transmission elsewhere, relative to the mean contribution of neighbourhoods.

$$\begin{aligned}
C_j(t) &:= \sum_{i \in \mathcal{N}(j)} \beta_{Mobility} \cdot w_t^{(j,i)} \cdot \log(\eta_{j,t-1}), \\
\overline{C_j} &:= \frac{1}{T} \sum_{k=1}^{T} C_j(t_k), \quad \overline{C} := \frac{1}{U} \sum_{j=1}^{U} \overline{C_j}, \quad \sigma(\overline{C}) := \sqrt{\frac{1}{U} \sum_{j=1}^{U} (\overline{C_j} - \overline{C})^2}, \\
\widetilde{C_j} &:= \frac{\overline{C_j} - \overline{C}}{\sigma(\overline{C})},
\end{aligned} \tag{5.2.11}$$

where T is the length of the rolling window (i.e., the 17-week period to which the model was fitted), U is the number (119) of neighbourhoods, $|\mathcal{N}(j)| = U - 1$ is the number of neighbours in $\mathcal{N}(j)$, and $w_t^{(j,i)}$ is the weight assigned for the proportion of outflows that leave neighbourhood j to visit neighbourhood i. Then, by tracking $\widetilde{C_j}$ in each rolling window, we measured how the relative contribution of a single neighbourhood j (to all other neighbourhoods) varies over time. We standardised per window to quantify *relative* contributions (source/sink) of neighbourhoods during different stages of dengue seasons (i.e., accounting for different epidemic magnitudes in different windows).

We assessed the roles of human mobility via these derived estimates (eqs (5.2.10) – (5.2.11)), alongside information criteria (ELPD LOO) and out-of-sample predictive performance for one-step-ahead forecasts using the Weighted Interval Score (WIS)[58] – a proper scoring rule that balances sharpness and calibration (see below).

5.2.4 Data-driven, deep learning models

We also developed data-driven, deep learning models (denoted as *DL models* in visualisations). The objective was to draw more robust conclusions by leaning on different approaches to modelling the effects of human mobility.

We implemented two foundational time series models (with and without human mobility effects), both based on TimeGPT – a transformer-based, foundational time series model that was pre-trained on more than 100 billion data points

from a range of domains[401,563]. In both models, we modelled the logarithm of incident cases y_{it} in neighbourhood i at time t as our response for two reasons; i) we constrained our model's predictions to be positive, and ii) we measured one-step-ahead probabilistic forecasts using a recommended logarithmic scale[213]. In only the mobility-informed deep learning model, we used a covariate of a mobility-weighted sum of other neighbourhoods' past cases $m_{i,t}$ (on a logarithmic scale), which entered via

$$m_{i,t} = \sum_{j \in \mathcal{N}(i)} \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}} \cdot \log(y_{j,t-1}), \tag{5.2.12}$$

where $\mathcal{N}(i)$ and $\mathrm{f}_{t-1}^{(j,i)}$ are defined as before. In both models (with and without the inclusion of human mobility effects), we further included the neighbourhood as a factor variable (i.e., a spatial random effect) and the neighbourhood's most recently observed cases $\log(y_{i,t-1})$ (on a logarithmic scale, i.e., similar to an autoregressive covariate). Unless otherwise stated, we use the mobility-informed model when discussing the deep learning model's results.

These foundational time series models facilitate probabilistic predictions (via quantile regression with conformal inference) in the form of predictive quantiles of the forecast/predictive distribution. Therefore, to ensure comparability when assessing models' probabilistic predictions, we used the WIS (via the scoringutils R package[560]), which is an approximation to the Continuous Ranked Probability Score (CRPS) for quantile-based forecasts, which we represented using 23 predictive quantiles, as in previous forecasting work[46,172].

A key challenge of any deep learning model is explainability[413]. To understand why our deep learning models made its individual predictions, we used SHAP (SHapley Additive exPlanations) values. These improve on feature importance metrics which can only measure overall importance to the model and cannot measure the directional influences of each feature. In contrast, SHAP values, based on a concept from game theory, measure the contribution of each feature/covariate to the model's individual predictions[418].

We denote $\mathrm{SHAP}_i^{(AR)}(t)$ and $\mathrm{SHAP}_i^{(Mob)}(t)$ as the respective SHAP values (i.e., estimated contributions) of the autoregressive influence, $\log(y_{i,t-1})$ and mobility-weighted inflows, m_{it} to the model's forecast of the logarithm of cases for neighbourhood i and time t. We can then derive an estimate for the relative importance of human mobility that is analogous to $r_i(t)$ (Eq (5.2.10)) from the Bayesian hierarchical model:

$$rSHAP_i(t) := \frac{\mathrm{SHAP}_i^{(Mob)}(t) - \log(P_{it})}{\mathrm{SHAP}_i^{(AR)}(t) - \log(P_{it})}, \tag{5.2.13}$$

where P_{it} is defined as before as the population offset (population divided by 100,000) needed to convert to the logarithm of DIR scale.

This formulation, eq. (5.2.13), again, can measure the estimated relative importance of human mobility in individual neighbourhoods with potentially different population sizes. Of course, we caveat our approach with the causal limitation that SHAP values measure feature contributions to the models' individual forecast. There is no explainability metric that can guarantee a correct causal result. Although our multi-model, triangulation-like approach[430] may mitigate this causal limitation, recent research has demonstrated causal misalignments when using SHAP values for deep learning models in epidemiological settings (even when the models are well-calibrated and highly predictive)[419]. In our setting, a SHAP value does not measure the contribution to observed case data nor to the true latent infection process. Furthermore, unlike our Bayesian hierarchical models which are generative and more mechanistic, the SHAP and rSHAP values of our deep learning models are not constrained by even a semi-mechanistic structure. For example, the SHAP value for the mobility-weighted influence $\mathrm{SHAP}_i^{(Mob)}(t)$ may be positive when $m_{it} = 0$ and could even be negative (when indicating that the base forecast should be decreased when m_{it} takes a particular, e.g., low, value). It is also possible that another unmodelled factor beyond mobility drives the difference between the SHAP values.

5.2.5 Evidence synthesis

Using both modelling approaches (Bayesian and deep learning), we generated one-step-ahead probabilistic forecasts of dengue cases on a logarithmic scale, and measured out-of-sample predictive performance of the mobility-informed models versus i) their analogous models with the mobility effect omitted and ii) a naive, baseline forecasting model[172]. Improvements over reference models, as measured by relative WIS, rWIS (WIS for a given model divided by WIS for an analogous model without mobility or baseline model), for different rolling time windows and neighbourhoods, allowed us to quantify the extent to which mobility-driven importations were important for generating the observed incidence data. For the Bayesian hierarchical model, we also measured improvements (upon the analogous model without mobility) using LOO estimates of the ELPD, thus measuring how much mobility effects improve (estimated) out-of-sample generalisability.

We interpret $r_i(t)$ and $rSHAP_i(t)$ from our Bayesian hierarchical model and deep learning model respectively as proxies for the relative importance of human mobility over time in driving epidemics. We compared values for $r_i(t)$ and $rSHAP_i(t)$ using Pearson's[586] and Spearman's correlation[587], categorisation into binary and four-level categories, and t-tests with p-values (Bonferroni-adjusted[588] when comparing multiple time series across neighbourhoods). Drawing comparison between these quantities ensured that we did not over-inflate the insights from a single modelling approach about the relative importance of human movement.

To understand the likely mechanisms behind the values for $r_i(t)$ and $rSHAP_i(t)$, we visualised relationships with factors such as epidemic stage (measured by the cumulative fraction of annual cases observed by a given week), year, and neighbourhood. We further developed simple generalised additive models (GAMs) for $r_i(t)$ and $rSHAP_i(t)$, consisting of a Gaussian likelihood with covariates of neighbourhood-specific smooth functions (thin-plate splines) for epidemic stage, annual temporal and neighbourhood-specific spatial random effects (to account for e.g., varying reporting rates), and the most recent observed cases $y_{i,t-1}$ on a logarithmic scale. We compared alternative GAM models using the minimised generalised cross-validation score (GCV)[553].

Due to its generative structure, the Bayesian hierarchical model permits estimates of how an individual neighbourhood j (via journeys of its infected residents/commuters) impact the epidemics in other neighbourhoods, as measured by $\widetilde{C_j}$. We quantified possible causes behind the values of $\widetilde{C_j}$ in different rolling windows using a similar procedure; visualisations and local contextualisation complemented by follow-up GAMs chosen based on their GCV. The candidate GAMs here are, by construction of $\widetilde{C_j}$, conceptually simpler as $\widetilde{C_j}$ is a standardised per-rolling-window quantity, and so temporal effects are likely less relevant.

After model fitting, using Pearson's correlation coefficients[586], we also performed a simple correlation analysis of the full DIR time series (2007–2015) for the 119 neighbourhoods to understand the spatial synchrony across dengue seasons. Finally, to better understand the structure of the mobility network, for each week, we used i) modularity from the Louvain algorithm[589] (a community detection algorithm that works by hierarchical clustering), ii) modularity from the Infomap algorithm[590] (a community detection algorithm for directed, weighted graphs), and iii) Shannon's entropy[591] of the mobility flow matrix. Higher values of modularity and lower values of entropy indicate strong clustering (i.e., low mixing) and high concentration of flows in particular locations.

5.3 Results

The 119 neighbourhoods of Fortaleza have different population sizes (ranging from 1,342 to 76,044, Figure 5.1 A) and therefore exhibit different human movement patterns (as approximated by the bus transportation data, Figure 5.1 B–D) and consequently, may have different daytime (or ambient) population sizes due to inflows or outflows caused by commuter journeys in the city. In general, pairwise total flows between neighbourhoods were similar throughout the year (Figure SI D.2), and there was a strong relationship (Pearson's correlation coefficient $r = 0.84$, p-value < 0.0001) between a

neighbourhoods' total inflows and total outflows.

Across the studied period (2007–2015), the most severe annual dengue seasons were observed in 2008, 2011, 2012, and 2015 (Figures 5.1 E and SI D.3 B), with peaks occurring almost across all neighbourhoods at similar times (Figures SI D.3 B, 5.3 A). We visualised (Figure SI D.3 A) and quantified this high spatial similarity of time series, reflective of epidemic synchrony, using pairwise Pearson's correlation coefficients[586] between individual neighbourhoods (Median r = 0.74, Q1 = 0.66, Q3 = 0.81).

On their own, these exploratory results suggest, but cannot confirm, roles for human mobility in driving dengue epidemics in Fortaleza. It is plausible that the highly synchronous dengue seasons across pairs of neighbourhoods (Figure SI D.3 A) are caused by single initial seeding events (as infectious hosts move from one neighbourhood to another) and subsequent similar epidemic trajectories unfold irrespective of human movement.

5.3.1 Mobility consistently drives urban dengue outbreaks

Across our statistical and deep learning models, including the effects of an estimated force of dengue importations (via weighted estimates of inflows, see Materials and methods) almost consistently improved model performance, as measured by probabilistic forecast performance metrics (rWIS, Figure 5.2 A-C) and information criteria (ELPD LOO, Figure 5.2 D). We observed these trends when we compared mobility-informed models to our baseline model which propagates the most recent observed cases as the next (median) forecast, and to analogous models which omitted the effects of the estimated force of dengue importations (Figures 5.2 D and 5.3 G). These trends were generally consistent i) across the 119 neighbourhoods, including in boundary neighbourhoods along the south and west of the city where our restriction to city-level mobility data could induce boundary effects (Figure 5.2 B) and ii) over time, including during both peaks and troughs of dengue cases (Figure 5.2 A and C). The only locations where we observed any predictive improvement from using the baseline model versus the mobility-informed Bayesian model were in Jacarecanga and Cristo Redentor; coastal/beach neighbourhoods which reported zero total dengue cases (Figure 5.1 E). We found largely similar outperformance for the mobility-informed deep learning model (in 116 of 119 locations, Figure 5.2 B). As a further model validation, the more generative (or semi-mechanistic) mobility-informed Bayesian hierarchical model consistently outperformed (within and across all dengue seasons) the corresponding data-driven deep learning model in terms of probabilistic predictive performance (WIS), suggesting greater out-of-sample model generalisability (Figure 5.1 A and C).

5.3.2 Relative importance of human mobility over time to individual neighbourhoods

How the relative importance of human mobility changes throughout dengue outbreaks in urban settings remains poorly understood. We hinted above that human movement could play its most important role at the early stages of a dengue season (when seeding events can occur as a result of between-neighbourhood movement). Thereafter, human mobility could play less important and potentially negligible roles in sustaining transmission, once dengue is circulating widely in neighbourhoods of the city.

Using our Bayesian model, eqs (5.2.1) – (5.2.9), we assessed relative importance of human mobility for each neighbourhood via our estimates of $r_i(t)$ – the ratio of (mobility-informed) case importations versus the neighbourhood's own autoregressive generation of new cases from past cases. Temporally, we estimated different trends in the relative importance of human mobility in different dengue seasons across the individual neighbourhoods (Figure 5.3 A-C, Figure SI D.5). While the rWIS values indicated spatially consistent (Figure 5.3 G) and temporally consistent (Figure 5.2 A, C, and D) outperformance of the mobility-informed model relative to the analogous model without human mobility, the estimates of $r_i(t)$ were spatially inconsistent within the annual dengue seasons (Figure SI D.5 C and Figure SI D.6). As $r_i(t)$ varied substantially across space and time, we used both a binary classification (Figure 5.3 C) and a four-level classification

(Figure 5.3 E) of $r_i(t)$ to further investigate the reasons/mechanisms behind variations in $r_i(t)$. Across all observations (all neighbourhoods and time points), the times of estimated greater relative importance of human mobility ($r_i(t)$) most frequently coincided with times near the midpoint (Figure 5.3 D) or epidemic peak (Figure SI D.5, SI D.6) of an annual dengue season. Despite these general observations, there is no distinct neighbourhood-level evidence of such systematic pattern.

To test the robustness of model-based conclusions, we assessed the relationship between the Bayesian and deep learning models' findings. First, we observed the same spatial and temporal outperformance of the predictions for the mobility-informed deep learning models, relative to analogous models without the mobility effect (as per rWIS, see Figure 5.2 A – C and Figure 5.3 G). We also compared the Bayesian-model-derived estimates $r_i(t)$ with those from the SHAP values for the mobility-weighted influence, $\text{SHAP}_i^{(Mob)}(t)$, on the deep learning model's predictions. The SHAP values indicate i) greater relative importance of the mobility-weighted sum of neighbours' cases, compared to the neighbourhood's own autoregressive process of case generation, at all times where cases are greater than zero, and ii) also that the mobility-weighted sum scales more drastically with the annual epidemics than the autoregressive process (Figure 5.2 E). This suggests a potential greater relative importance of human mobility (versus the autoregressive process) for the persistence of dengue in these urban environments.

Of course, $r_i(t)$ and the SHAP mobility influence estimates are not directly comparable, as the SHAP value scales with the magnitude of the data (e.g., Figure 5.2 E), unlike our Bayesian hierarchical model (which models DIRs using population offsets). Therefore, we used the relative SHAP quantity, $r\text{SHAP}_i(t)$, computed as the ratio between the mobility-weighted influence (to DIR) and autoregressive influence SHAP values (see Materials and methods). We found that there was no one-to-one, linear relationship between $r_i(t)$ and $r\text{SHAP}_i(t)$ (Figure SI D.5), perhaps due to the lack of a mechanistic or generative structure in our deep learning model which could constrain the range of plausible values for $r\text{SHAP}_i(t)$. However, times with greater relative importance of human mobility from the Bayesian model were consistently associated with higher SHAP values for the mobility effect (Figures 5.3 H, SI D.5, SI D.6), particularly in years with larger dengue outbreaks. Using t-tests, we formally tested this relationship, between $r_i(t)$ and $r\text{SHAP}_i(t)$ across all observations and for each of the 119 neighbourhoods individually (see Figure 5.3 F for Bonferroni-adjusted[588] p-values). For both models, we estimated similar trends for the timings of maximal importance of human mobility (middle stages of season, Figures 5.3 E-F, SI D.5, and SI D.6).

The maximal relative importance of mobility was most often observed at the midpoint of the annual dengue season (Figure 5.3 E-F), and the relative importance values generally tracked the seasonal epidemic curve (Figure SI D.6), yet this estimated pattern varied across seasons (Figure SI D.7) and across neighbourhoods (Figures SI D.8 – SI D.9). Alongside the consistent high values of $r_i(t)$ (above 1), these results collectively suggest a potential longer-term trend of human-mobility-driven spread (in urban environments) being consistently more important than internal case generation across all stages of the dengue season, and likely most important in early-to-mid stages of the dengue season, both in more and less severe dengue seasons. Mechanistically, this implies that human-mobility-driven case importations/amplifications may account for sustained circulation of (reported) dengue cases, particularly at seasonal peaks (Figure SI D.6).

To further investigate the plausibility of these findings, we explored the human movement patterns using various mixing metrics, such as modularity (from the Louvain and Infomap algorithms) and Shannon's entropy for the mobility matrix (Figure SI D.10). We estimated consistently low modularity (and the Infomap algorithm did not detect any communities) which indicated a well-mixed mobility network in Fortaleza, at least at a neighbourhood level of analysis. Similarly, the entropy was high across all weeks (Median = 7.96, minimum = 7.87, and maximum = 8.06), relative to the maximum achievable of $\log(119^2)$ = 9.56. This high entropy indicates more mixing and more evenly spread flows throughout Fortaleza. Collectively, the results indicate that high mixing may have generated a consistently high importance of human movement and widespread epidemics.

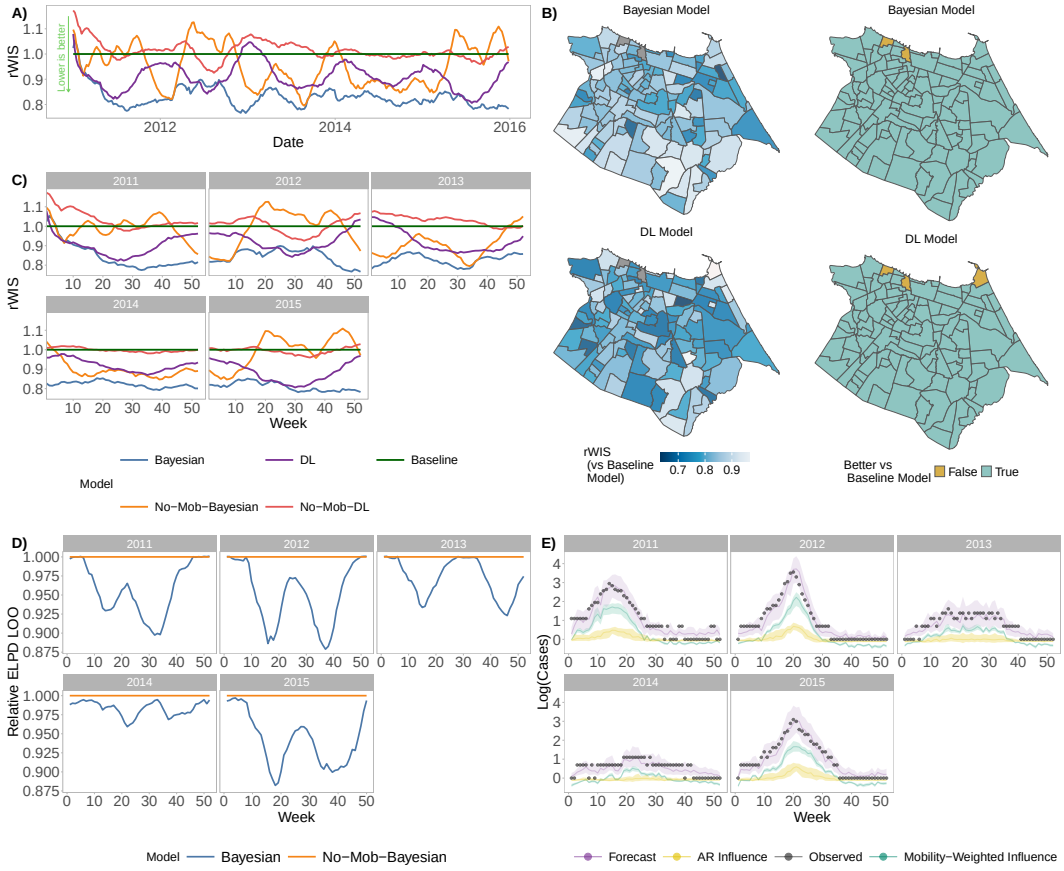


Figure 5.2: **Capturing human mobility effects improves model performance:** In approximate four-month trailing windows (17 weeks) of the testing period (2011-2015 inclusive), we visualise model performance for each of our four models across Bayesian hierarchical and deep learning (DL) approaches and compare performance to an epidemiologically naive baseline model (see Materials and methods). Any date shown represents the maximum date within the trailing window. **A)** Relative weighted interval score (rWIS) over time for each of the four models, where rWIS was calculated (relative to a baseline model, see Section 5.2.5) out of sample across all neighbourhoods. Lower values of the rWIS indicate superior predictive performance for out-of-sample case data (on a logarithmic scale). **B)** rWIS values for each of the 119 neighbourhoods across all trailing windows and a binary classification of whether the Bayesian/DL model outperformed the baseline (i.e., rWIS < 1). The grey shaded neighbourhoods on the north coastline are Jacarecanga and Cristo Redentor; neighbourhoods which reported zero total dengue cases across all seasons. **C)** Analogous to A), rWIS over time but using separate facets to capture model performance within each dengue season. **D)** Information criteria results for mobility-informed Bayesian hierarchical model. Higher ELPD LOO values indicate superior model generalisability. However, for visualisation of relative model performance over time (irrespective of the magnitude of observed data), we derived relative ELPD LOO values by dividing the ELPD LOO values for the Bayesian model by those for the analogous model without mobility (No-Mob-Bayesian). This was possible, as all ELPD LOO values were negative (across all windows and both models). Therefore lower values of this relative ELPD LOO indicate superior estimated model generalisability. **E)** Interquartile ranges of observed and forecasted cases (on a logarithmic scale) alongside the SHAP contributions from the neighbourhood-level autoregressive (AR) process and the mobility-weighted influence of the neighbours. Higher values of the SHAP values indicate greater contributions to the forecasted value.

5.3.3 Space-varying human-mobility-driven influences of neighbourhoods

Different neighbourhoods in a city can play different roles in daily life (e.g., serving as commuter hubs, retail epicentres, or residential areas), and therefore play different roles in seeding and/or sustaining the dengue epidemics of other neighbourhoods. We used our Bayesian hierarchical model, eqs (5.2.1) – (5.2.9), in rolling windows alongside the derived quantity $\widetilde{C_j}$, eq. (5.2.11), to investigate which neighbourhoods of Fortaleza were key contributors (i.e., sources) to the dengue outbreaks in *other* neighbourhoods, and to determine whether these neighbourhoods played consistent source

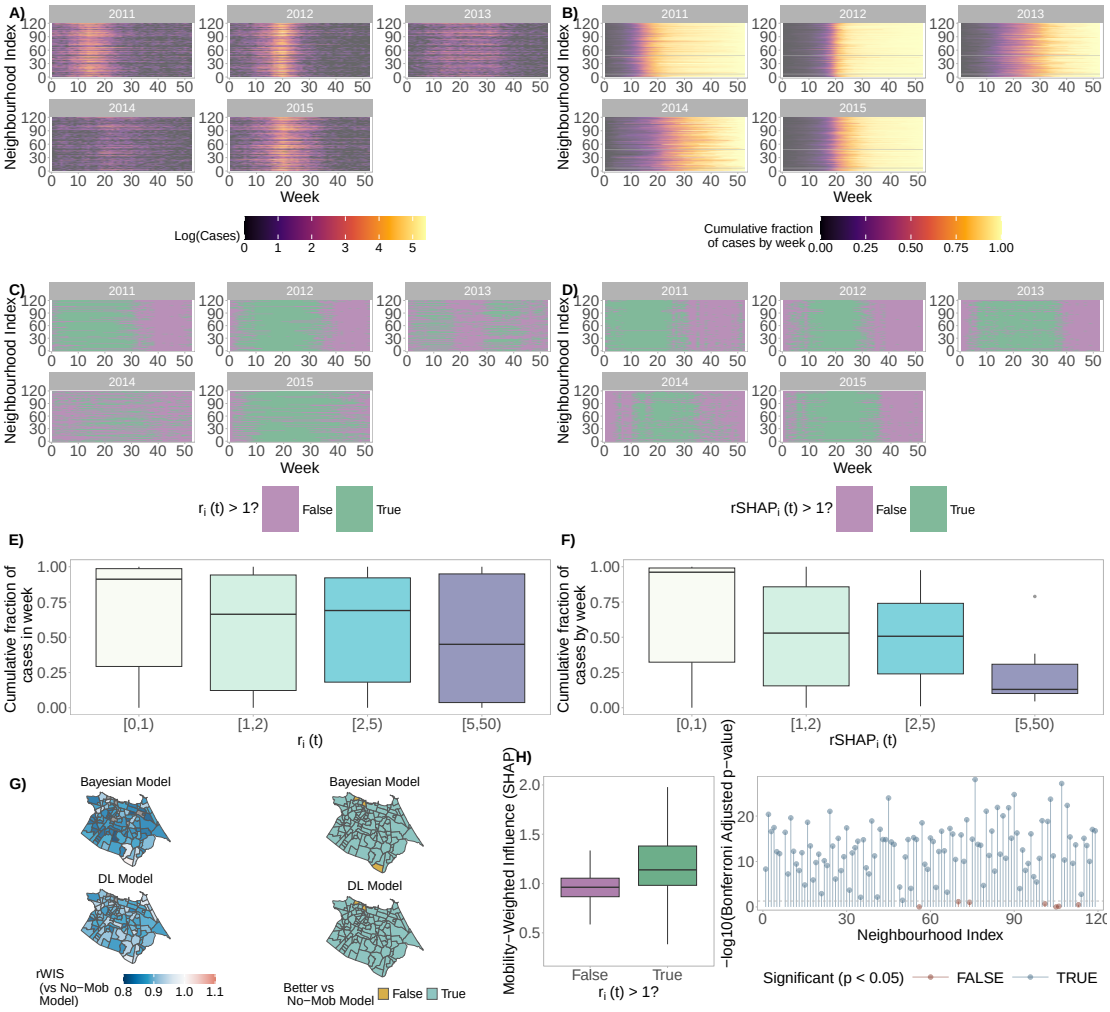


Figure 5.3: **The time-varying relative importance of human mobility:** Across 2011–2015 and the 119 neighbourhoods, we visualise the estimated relative importance of human-movement-based importations of dengue cases, versus the neighbourhoods' own internal generation of new dengue cases. **A)** and **B)** visualise the observed cases (on a logarithmic scale) and cumulative fraction of annual cases observed for individual weeks across 2011-2015. Neighbourhoods are ordered in ascending order by their total annual number of flows (inflows and outflows). **C)** and **D** show the estimated relative importance from our Bayesian hierarchical and deep learning models respectively, where values of $r_i(t)$ and $rSHAP_i(t)$ greater than one indicate greater relative importance of the case importations (see Materials and methods). Figure SI D.5 C-D is the analogous continuous version of these estimates of $r_i(t)$ and $rSHAP_i(t)$. **E)** and **F)** visualise the relationship between dengue season stage (measured by the cumulative fraction of cases) versus the values of $r_i(t)$ and $rSHAP_i(t)$. **E)** Across the 119 neighbourhoods, the relative predictive performance of the Bayesian and deep learning model is visualised, using the relative weighted interval score (rWIS) compared against the corresponding models without a mobility effect. **G)** and **H)** capture the relationship between the model-based relative importance estimates $rSHAP_i(t)$ and a binary version of $r_i(t)$ (= 1 if > 1) using a boxplot and visualisation of the Bonferroni-adjusted p-values[588] from a one-sided t-test of whether $rSHAP_i(t)$ were significantly greater when $r_i(t) > 1$. The box and whiskers of each boxplot cover the interquartile range.

roles within each season and across all seasons. As $\widetilde{C_j}$ was standardised per window (see Materials and methods), we could compare the neighbourhoods' (relative) contributions at different stages of dengue seasons. This comparison of neighbourhoods' contributions over time differs from above, where we were concerned with the estimated time-varying relative importance of human mobility to each neighbourhood's dengue season, as here we are focused on source-sink dynamics. $\widetilde{C_j}$ does not represent a neighbourhood's overall importance for the public health burden in dengue, but rather the human-mobility-driven relative influence of neighbourhood j on the epidemic dynamics in all other neighbourhoods.

We found a strong, spatially consistent pattern of high-mobility neighbourhoods consistently being the greatest contributors, in rolling windows, to the epidemics of other neighbourhoods. This pattern was replicated across different stages of individual dengue seasons (Figure 5.4 A), and across all dengue seasons (Figure 5.4 B–C). There was generally little variability in the standardised contributions of individual neighbourhoods across the 258 rolling windows (Figures 5.4 C and SI D.11). Indeed, a GAM for median $\widetilde{C_j}$ values with a very simple Gaussian linear model structure, neighbourhood-specific random effect, and intercept (i.e., no temporal effects) explained 84% of the variation in $\widetilde{C_j}$ (R^2).

Spatially, while intuitive, it is not guaranteed that neighbourhoods with high mobility flows are consistently the greatest contributors. The contributions of generally high-mobility neighbourhoods could be influenced by i) weekly variability in pairwise mobility patterns and/or ii) the stage of the epidemic (as a high-mobility neighbourhood could, in some windows, have relatively few infectious hosts, relative to other neighbourhoods). In our case, both factors, i) and ii), are apparently not applicable as we found temporally homogeneous mobility patterns (Figure SI D.2) and high epidemic synchrony in the city (Figures SI D.3 and SI D.4). These findings are limited by our dependence on bus transportation data only (see Discussion). For example, the neighbourhood of Siqueira had the greatest inflows (Figure 5.1 B), outflows (Figure 5.1 C), and estimated contributions to other neighbourhoods (Figure 5.4 A–C), and is home to a bus terminal, which could inflate this neighbourhood's estimated influence on other neighbourhoods.

The only exception to the high relative influence of high-mobility neighbourhoods was during the "peak" of the low-incidence 2013 season, when their estimated relative contribution to other neighbourhoods was lower than lower-mobility neighbourhoods. This phenomenon may be partially explained by the highest cases during this period not being in the highest-mobility neighbourhoods (Figure SI D.12) and the lower synchrony of the outbreaks in the low-incidence season of 2013 (Figure SI D.4, median pairwise Pearson's correlation coefficient[586] r = 0.36, Q1 = 0.22, Q3 = 0.49), yet the high-mobility neighbourhoods did not have the lowest cases during this wave (Figure SI D.12) and this spatial mobility influence trend was not replicated in the next lower-incidence season of 2014. This phenomenon could, of course, be a modelling limitation as the ELPD LOO indicates poorer generalisability, compared to a model without mobility, during this wave (Figure 5.2 D)

5.4 Discussion

Human movement is widely understood as an important driver of many respiratory, vector-borne, water-borne, sexually transmitted, and other infectious diseases[224,225,226], shaping the propagation of epidemic dynamics across space and time. However, for dengue epidemics in large cities, it was less clear how i) the relative importance of human mobility and ii) the human-mobility-driven source-sink dynamics of different neighbourhoods, vary within and across individual dengue seasons. Here, we provided multi-model methods to investigate these questions, illustrating our application in high spatial resolution for urban dengue outbreaks in Fortaleza, Brazil.

Understanding *when* human mobility is most important in an outbreak has practical consequences for public health policy. This can inform the optimal timing for public health interventions (e.g., vector control strategies) for endemic, epidemic, and pandemic pathogens. Understanding *where* human movement is seeding, expanding, and/or sustaining outbreaks of other geographies can enable more targeted and cost-effective interventions. For example, in our study, while we estimated that human mobility was consistently important across neighbourhoods throughout a dengue season, we identified a high relative importance i) of human mobility at the middle stages and peaks of dengue seasons and ii) of particular neighbourhoods to geographic spread in each dengue season. Targeting vector control interventions to reduce human-vector contact within such times and neighbourhoods could have an outsized, positive impact on bringing the epidemic under control.

Beyond public health interventions, such insights could also be used to inform the design of human movement data col-

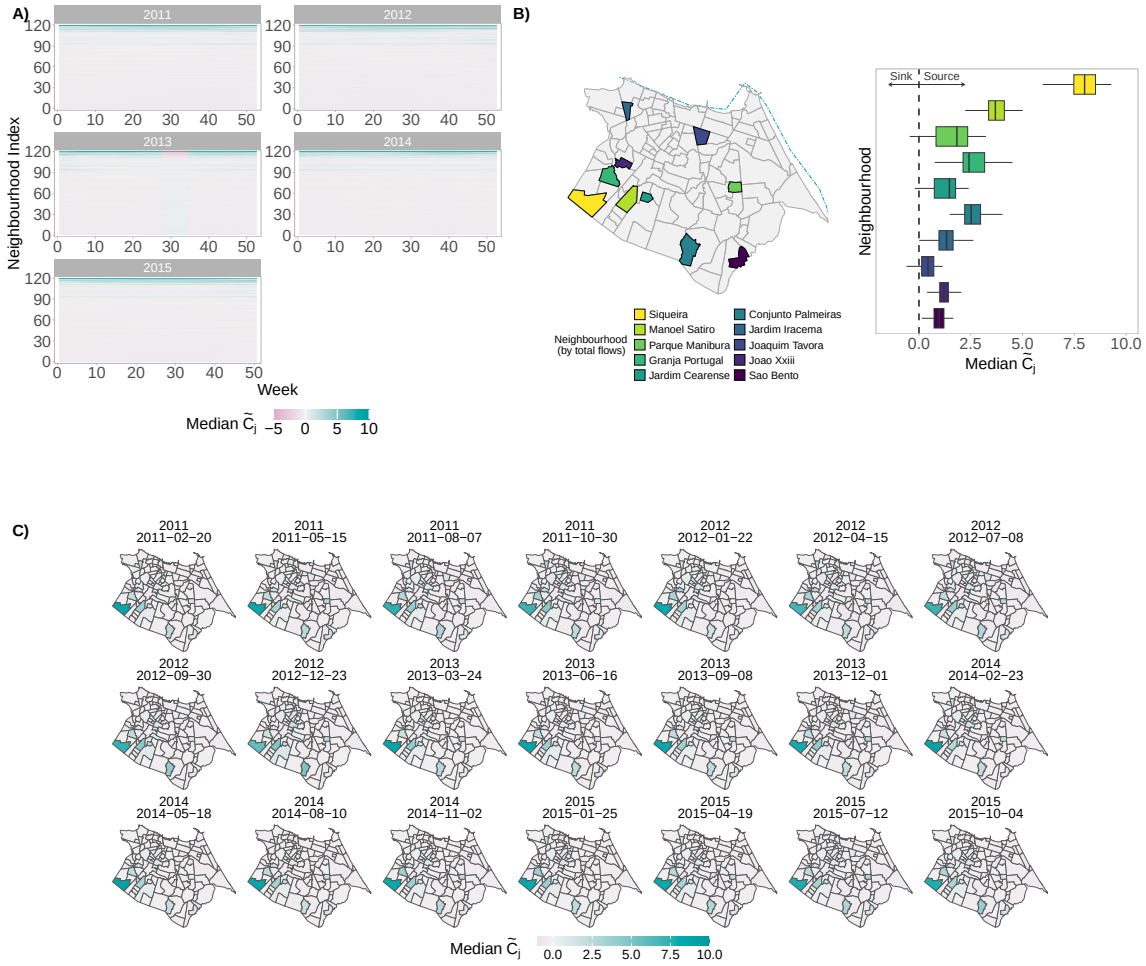


Figure 5.4: **Space-varying source-sink dynamics within and across dengue seasons:** Estimated relative (standardised) contributions of individual neighbourhoods to the epidemics of all other neighbourhoods, where estimates are based on the fits of our Bayesian hierarchical model, eqs (5.2.1) – (5.2.9), and the derived quantity $\widetilde{C_j}$, eq. (5.2.11). The model was fitted in rolling windows of length 17 weeks to allow for sufficient data for model training, while allowing for potentially time-varying effects. **A)** and **C)** are the median $\widetilde{C_j}$ for different neighbourhoods in different rolling windows across 2011–2015, where the week/date visualised represents the maximum date of the rolling window. Neighbourhoods in A) are ordered in ascending order from 1 to 119 by their total annual number of flows (inflows and outflows). Note that in **C)**, we visualised results for only every 12th rolling window (i.e., every 12 weeks) for legibility, and this results in different limits to the legend colours (compared to A). **B)** is a boxplot (right) for the neighbourhoods with the top ten total annual number of flows (left), where for each neighbourhood, all rolling window median $\widetilde{C_j}$ values are summarised. The box and whiskers cover the interquartile range.

lection strategies. Routine mobility data collection and sharing, particularly in low- and middle-income countries (LMICs), is challenging[238] (due to economic, political, and logistical barriers) and often infeasible, therefore suggesting that more cost-effective and targeted data collection is needed. Although our focus here has not been on forecasting (see[46,86,162,557,592]) nor invasion dynamics (see[227,355,593]), we have demonstrated consistently strong (out-of-sample) predictive ability of our models via the enhancement of spatiotemporal predictions of dengue incidence by using human mobility data. Therefore, incorporating the spatial structure induced by human mobility in invasion and forecasting models is likely important for more accurate forecasts that can be used for short-term outbreak response and longer-term seasonal planning.

Our results complement previous work using similar and other data streams to investigate other questions about the influences of human mobility on shaping the spread of infectious diseases. In addition to public transportation data, researchers have used human movement data from air and water travel[273], population censuses[594], and mobile

phones[269,275] to investigate questions of disease spread on subnational, national, and global scales. Future work may look at understanding how best to integrate these heterogeneous mobility data sources with our methods and analyses (potentially extending research questions to roles of long-distance human movements) as well as the net information gains (including corresponding costs and potential data biases) from such additional data collection. These net information gains will depend on the setting-specific research questions (e.g., early warning detection, nowcasting, or forecasting) and the time and region, but quantifying the information gains is crucial for encouraging and guiding data collection. This could be a stepping stone towards addressing our aforementioned barriers caused by the lack of subnational human mobility data, particularly in resource-limited LMICs[238,266]. Understanding data information gains is also important, as different model-based proxies[238] and mobility data streams[266] can induce biases in the modelling estimates that are ultimately used to inform public health authorities, although there are recent advances in the accuracy of mobility data proxies (e.g., gravity-like models[595]) and integration of heterogeneous mobility data streams[267].

The consistently high relative importance of human mobility (and peaks coinciding with seasonal epidemic peaks) and of specific source neighbourhoods makes intuitive sense within the context of a large, highly urbanised, well-mixed, and commuter-centric city such as Fortaleza. Mechanistically, once the seasonal urban dengue outbreak becomes established, the movement of humans and thus the provision of infectious hosts apparently allows for the epidemic surges (i.e., expansion) and sustained spread across neighbourhoods. In previous work, arbovirus risk (for dengue, Zika, and chikungunya) in Fortaleza (albeit focused on 2015 onwards) had estimated spatial clusters of cases[184], including for clusters of neighbourhoods over longer distances (up to 6.8 km). The authors hypothesised that human mobility likely allowed infected individuals to regularly import viruses to distant areas of Fortaleza. Similarly, infestation surveys, which estimate *Ae. aegypti* density, indicate a possibly high level of vector adaptation across Fortaleza[579]. Interpreted alongside a consistent role for human movement, this suggests that even in less suitable climatic conditions (e.g., dry season), the mosquito vector could be present near homogeneously across space and time. This example underscores the importance of understanding the history of dengue transmission when applying our framework and interpreting the results in other settings. In recently affected cities (e.g., Buenos Aires, Argentina), the likely fluctuations in mosquito vector density (as the effects of climate change materialise[355,356]) could modulate mobility effects and yield very different influences of human movement on urban dengue epidemic dynamics.

Meanwhile, for COVID-19, Aguilar et al.[596] compared hierarchical cities with sprawled/well-mixed cities, and found that epidemic spread was less rapid in sprawled cities. Interpreting our results, albeit for a different infectious disease, alongside these findings, we suspect that the high mixing across the city (Figure SI D.10) may have supported the relatively long dengue seasons frequently observed in Fortaleza (Figure SI D.3 B). This may explain why we estimated a high relative importance of mobility in multiple individual dengue seasons. Further work could investigate whether these results generalise in other environments – different mobility patterns, different spatial scales of cities, and different socio-environmental settings.

Our findings align with research that documented consistently superior overall predictive performance of mobility-informed models in Fortaleza[86,233], other cities (e.g., 12 cities in Guangzhou province, China[235]), and at national levels (e.g., in Pakistan[85] and Vietnam[79]). Our research also builds on such research by disentangling the relative importance and influences of human movement for different stages of the epidemic and for different neighbourhoods. For future studies, our unitless measures of the time-varying relative importance of human mobility ($r_i(t)$, eq. (5.2.10), and $r\text{SHAP}_i(t)$, eq. (5.2.13)) and of different neighbourhoods ($\widetilde{C_j}$, eq. (5.2.11)) could be useful. These could be paired with new formulae for other unitless measures of epidemic properties such as time-varying reproduction numbers $R(t)$ (e.g., for dengue[48,375]) to further quantify how the relative importance of human mobility and the influence of different neighbourhoods evolve with varying epidemic strength (as measured by $R(t)$, the number of new infections generated by a typical individual, see[135]).

The models that we develop are new for modelling the time- and space-varying roles and effects of human movement

on dengue epidemics, yet the models remain conceptually simple. The key advance is the ability to derive mathematical quantities (together with robust uncertainties) for the importance and influence of human mobility from a generative, Bayesian model, and to blend the results for these quantities with those using our derived formulae for game-theoretic quantities (i.e., SHAP and rSHAP values). This multi-model procedure for extracting spatiotemporal insights (by testing theories and synthesising evidence across disciplines) enhances robustness of causal interpretations[430], and our conceptual advances are generalisable to other settings (e.g., other infectious diseases) and other research questions of relative importance (e.g., climatic versus human influences).

Our analysis has several data limitations and there are many future avenues of research. First, we used weekly bus transportation data from 2015 to represent movements for all years. Although there is potential for representation bias from any form of mobility data[266] and we advocate for greater availability of more regular data, previous analyses have shown that these data are representative for a wide range of explanatory and predictive research questions (e.g., crime rates, dengue epidemics and police allocation)[86,581,582,583]. Similarly, there were no substantial time variations in flows throughout the weeks of 2015 (Figure SI D.2) and our mobility-informed models consistently enhanced performance across information and predictive performance criteria, hinting at the generalisability of these data (beyond 2015) and at the potential value of intermittent collection and sharing of such data. However, the aggregated bus data only capture shorter-distance movements (as opposed to importations from inward travel to the city) are not linked to demographic or social identifiers, nor specific disease status, and future work could explore the additional value or differences in results from integrating other individual-level human movement indicators (e.g., mobile phone data) and long-distance movement indicators (e.g., flight data). Such data sources may also have greater variability throughout the dengue season and provide valuable additional information for understanding the dynamics of invasion and geographical spread. Our models did not explicitly account for within-neighbourhood mobility, as bus transportation data were more suitable (within-neighbourhood bus trips were rare) for our studied research question – comparing imported cases with those generated within neighbourhoods, which means that within-neighbourhood cases can therefore account for human, biological, and environmental effects. Future work, likely relying on shorter-range movement data (compared to bus transportation data), could explore the relative importance of within-neighbourhood mobility over space and time. This would be instructive for better understanding of local community transmission in large informal settlements, which could be supported by the movement of infectious individuals on foot or motorcycles, coupled with the spread of mosquito breeding sites.

Further data limitations include our usage of observed dengue case data which are an imperfect measure of true dengue infections. The case data may be sensitive to reporting delays and/or underreporting in time or space (e.g., due to health centre availability or reporting capacity/propensity), which we attempted to proxy by using neighbourhood-specific random effects. These issues may also be exacerbated by the high spatial resolution of our study, yet this is balanced against the benefits of i) avoiding potential spatial aggregation artefacts/effects and ii) providing high-resolution insights for targeted policy questions.

Meanwhile, we did not have access to serotype or seroprevalence data. This may limit the robustness of our illustrative application as all four DENV serotypes have circulated in Fortaleza and 2011 and 2012 outbreaks were caused by different serotypes (the reintroduction and introduction of DENV-1 and DENV-4, respectively), which was the first time that a major Brazilian city experienced a serotype switch in two consecutive annual dengue epidemics. In 2015, DENV-1 was dominant once again. Overall, our results are probably not confounded by more recent introductions of Zika, and chikungunya viruses (which would induce misclassifications, yet these were almost simultaneously first reported in 2015)[184]. We also used only the one-week-lagged cases and did not model the mosquito population directly, and therefore we only approximated the true causal dynamical processes (e.g., biological mechanisms) of DENV transmission. Further causal limitations include the limitations of explainability metrics (such as SHAP values) for deep learning models when disentangling causal effects[413,419] (see Materials and methods). In future work, contingent on the availability of new data

streams (seroprevalence, genomic, higher spatial resolution climate, or mosquito catch data), researchers could consider alternative networked mechanistic (or biomathematical) and/or deep learning models. These models could enhance the causal realism with explicit time- and space-varying properties (e.g., generation time and serial interval distributions[48,375]). Similarly, models could incorporate additional human influences (e.g., population immunity[438], access to clean drinking water[78,84] and circulating serotypes[79,574]), climatic drivers (e.g., interactions with precipitation, drought, humidity, or temperature[45,78]) and/or mosquito factors (e.g., mosquito density[82,88]).

In conclusion, our theoretical advances (evidence synthesis across multiple models) and findings (influences of human movement) can be a foundation for i) future studies that improve our understanding of the complex processes of dengue epidemics in different settings and ii) future public health policy that targets limited resources for dengue epidemic control.

Chapter 6

Reproduction numbers for vector-borne diseases

Cathal Mills, Tarek Alrefae, William S. Hart, Moritz U. G. Kraemer, Kris V. Parag, Robin N. Thompson, Christl A. Donnelly, and Ben Lambert. Renewal equations for vector-borne diseases, September 2024. URL `http://arxiv.org/abs/2409.18726`. arXiv:2409.18726 [q-bio]

This chapter presents theoretical foundations for estimating reproduction numbers for vector-borne diseases[48]. The manuscript is currently in the editorial review stage at *Methods in Ecology and Evolution* following the completion of minor revisions. The manuscript introduces an age-structured modelling framework of PDEs across mosquitoes and humans, which allows us to derive renewal equations to describe how past human infections generate new human infections.

Abstract

During infectious disease outbreaks, estimates of time-varying pathogen transmissibility, such as the instantaneous reproduction number $R(t)$ or instantaneous growth rate r_t, are used to inform decision-making by public health authorities. For directly transmitted infectious diseases, the renewal equation framework is a widely used method for measuring time-varying transmissibility. The framework uses information on the typical time elapsing between an infection and the offspring infections (quantified by the generation time distribution), and $R(t)$, to describe the rate at which currently infected individuals generate new infections, if conditions affecting human transmission were to remain the same as at time t. For diseases with transmission cycles involving hosts and vectors, however, renewal equation models have been far less used. This is likely due to difficulties in mechanistically defining generation times that can capture the complexity of multi-stage, human-vector relationships. Here, using dengue as an example, we provide general renewal equations that are derived from first principles using age-structured systems of coupled partial differential equations across human and vector sub-populations. Our framework tracks the multi-stage transmission cycle over calendar time and across stage-specific ages, resulting in governing renewal equations that quantify how the rate at which new infections are generated from existing infections depends on stage-specific processes. In a numerical application to real-world temperature data, we show how the generation time distribution depends on both current and historical conditions. The framework provides a foundation on which to base inferential frameworks for estimating $R(t)$ and r_t for infectious diseases with multiple stages in the transmission cycle.

Author contributions

- Conceptualization: **C.M**, B.L
- Methodology: **C.M**, B.L;
- Data curation: **C.M**.
- Data visualisation: **C.M**.
- Writing original draft: **C.M**; T.A, W.S.H, M.U.G.K, K.V.P, R.N.T, C.A.D, B.L.
- Writing final draft: **C.M**; T.A, W.S.H, M.U.G.K, K.V.P, R.N.T, C.A.D, B.L.

6.1 Introduction

In infectious disease epidemiology, renewal equations can be used to describe how new infections are generated from individuals infected in the past. The time period that elapses between a previous infection and a new infection caused by it is called the generation time, and the variation in this time period across individuals is captured by the generation time distribution[597]. The number of infections, on average, that a given infected human would generate (should conditions affecting human transmission remain the same) throughout the course of their infection is known as the instantaneous reproduction number, $R(t)$[137].

For directly transmitted (human-to-human) infectious diseases, renewal equation frameworks have found widespread use for estimating $R(t)$. The popularity is likely in part due to the relatively few assumptions needed to produce $R(t)$ estimates compared to other frameworks such as compartmental models. By fitting renewal equation models to time series of infection counts, or more commonly, case counts, it is possible to estimate $R(t)$, and these models form the basis for many popular approaches and software used to infer changes in transmission throughout outbreaks[137,139,140,148,149]. A renewal equation model for a directly transmitted infectious disease may be described by

$$i(t) = R(t) \int_0^\infty i(t-\tau) w(\tau) d\tau, \tag{6.1.1}$$

where $i(t)$ is the number of new infections generated at time t, $R(t)$ is the instantaneous reproduction number at time t, and $w(\tau)$ is the probability density that describes the generation time distribution, resulting in $\int_0^\infty w(\tau) = 1$. Eq. (6.1.1) intuitively gives a process by which previous infections beget new infections with some time delay, and if $R(t) > 1$, the epidemic would grow in the long term if conditions affecting human transmission were to remain the same as at time t. Consequently, the threshold of $R(t) > 1$ coincides with an instantaneous epidemic growth rate $r_t > 0$, where r_t is the rate of change of the log infection incidence in calendar time.

The renewal equation is most often implemented in discrete time to align with the temporal resolution of surveillance data[598]. Renewal models, as used for inference, also often include stochasticity (e.g.,[152]). Misspecification of the generation time distribution can be a large potential source of error in this estimation of $R(t)$[113,134,142,145], and given the popularity of $R(t)$ for informing public health policy and decision-making[118], the violation of the assumption of time-invariant generation times clearly has practical, real-world consequences.

For infectious diseases with a single population (equivalently, a single transmission stage) and a static generation time distribution, continuous renewal equations (such as eq. (6.1.1)) can be derived from a partial differential equation (PDE) representation of the infection-age-structured model of the infected population; this was first derived by Kermack and McKendrick[330] [374,599]. The renewal equation can also be used to estimate so-called state reproduction numbers, starting from an age-structured system with multiple host populations[600]. We provide a short description of these methods in Appendix E.1, and for a more complete primer on renewal equations for directly transmitted diseases, see Bouros et al.[370].

For vector-borne diseases, renewal equations are less commonly used as i) it is unclear whether the popular renewal equation can generalise to the multi-stage transmission cycle across various host and vector sub-populations (e.g., four stages for dengue across mosquito and human populations, Figure 6.1), and ii) the common assumption in renewal equations (as in eq. (6.1.1)) of a static generation time distribution is often unrealistic for climate-sensitive, vector-borne diseases. Under laboratory conditions, several mosquito-viral traits, such as the extrinsic incubation period, the rate of mosquito blood feeding, and mosquito density, have shown to be influenced by climate factors such as temperature[80,82,83]. Also, the time taken for an infectious (or susceptible) mosquito to bite a susceptible (infectious) human will depend on susceptible human (mosquito) population levels.

Existing approaches have used temperature data to parameterise temperature-dependent, time-varying generation times to estimate basic (R_0) and instantaneous ($R(t)$) reproduction numbers. These approaches often involve a convolution of probability densities for the stages of the dengue transmission cycle, Ross-Macdonald assumptions for mosquito-borne pathogens, and/or parametric distributions for temperature-dependent generation times[83,121,156,325,375,601]. For example, one such existing approach is Siraj et al. [156] which provided formulae for a time-varying generation time distribution based on current temperatures. Such approaches, however, do not provide a mechanistically derived generation time distribution – as the relative contributions of previously infected humans and mosquitoes to new human infections. Another approach for vector-borne diseases has used renewal equations, albeit with a static generation time distribution[602].

Here, we adopt a different approach by deriving renewal equations for a vector-borne disease starting from an age-structured system of coupled PDEs. To the best of our knowledge, this is the first derivation of renewal equations from first principles for a vector-borne disease. In doing so, we derive a mechanistically based generation time distribution which informs estimates of $R(t)$ from our renewal equation, where $R(t)$ retains the same interpretation (as for directly transmitted diseases) – the number of infections, on average, that a given infected human would generate (should conditions affecting human transmission at time t remain the same) throughout the course of their infection. Throughout, we use dengue as a case study to motivate and develop our renewal equations, yet our approach can generalise to other mosquito-borne diseases (e.g., Zika, malaria, or yellow fever).

6.2 Methods and results

6.2.1 Infection-age-structured human and mosquito populations

Modelling framework

To illustrate the calculation steps for our full model of the dengue transmission cycle (Section 6.2.2), we start with a simplified abstraction of the dengue transmission cycle. We do not explicitly track the calendar times when a human (mosquito) becomes infectious after the viral infection event (i.e., after a mosquito bites a human). Our age-structured framework (further details of our framework are in Appendix E.2) describes both the infected mosquito and human populations using continuity equations: the McKendrick–von Foerster equation.[368,603].

Our system of age-structured populations is described by the following conservation equations:

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a_E} = 0, \tag{6.2.1}$$

$$\frac{\partial M}{\partial t} + \frac{\partial M}{\partial a_M} = -\mu M(t, a_M), \tag{6.2.2}$$

where t denotes calendar time, $E \geq 0$ and $M \geq 0$ denote the density of the infected human and mosquito populations respectively with corresponding ages $a_E \geq 0$ and $a_M \geq 0$, and $\mu > 0$ is the (fixed) mortality rate in the infectious mosquito population. The variables a_E and a_M denote the time since entering the stage (*not* the chronological or biological age

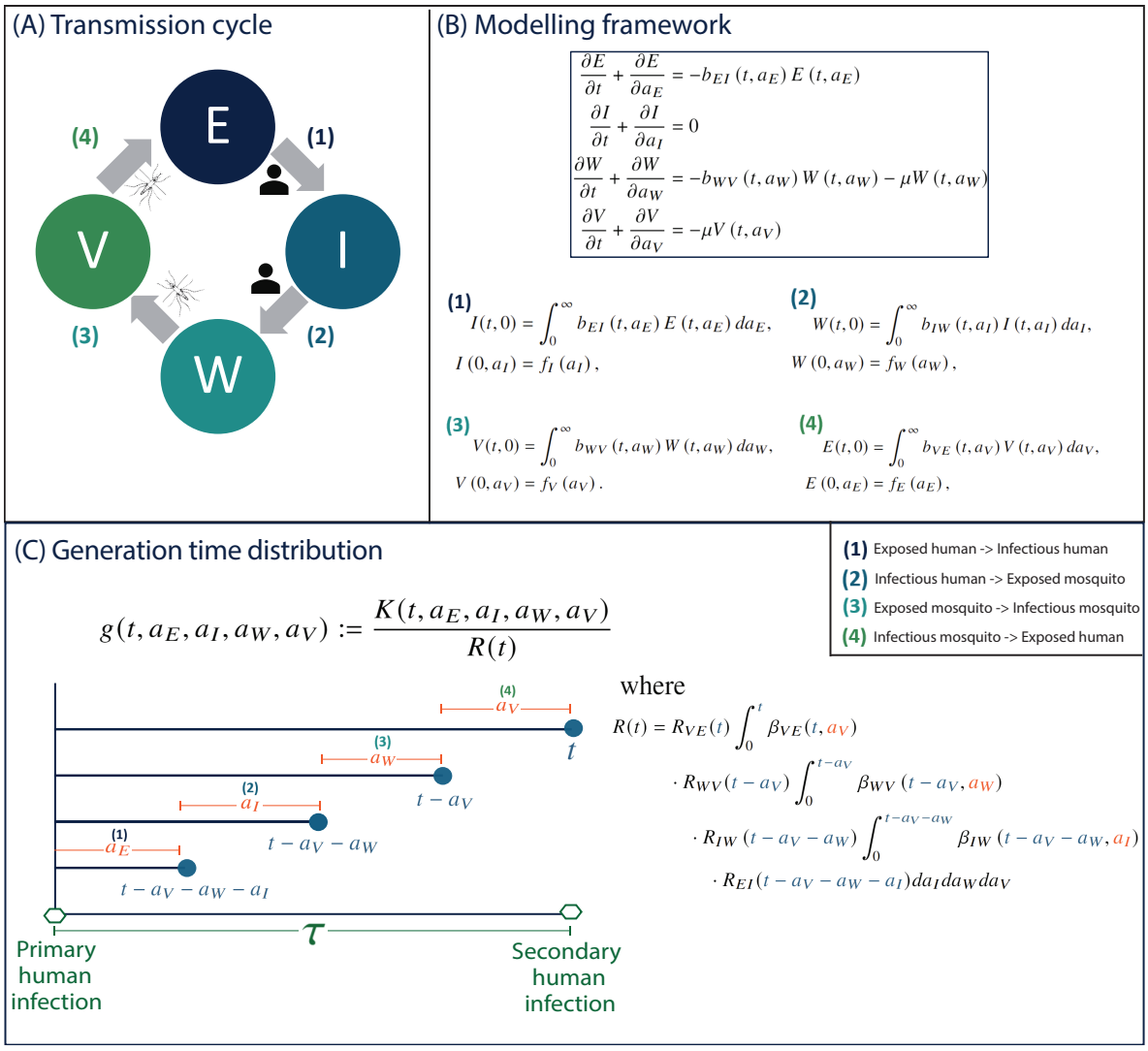


Figure 6.1: **Schematic of key concepts used to estimate $R(t)$ for mosquito-borne diseases**. A) The four-stage transmission cycle for dengue – from exposed humans (E) to infectious humans (I) to exposed mosquitoes (W) to infectious mosquitoes (V). The boundary conditions (numbered) of our modelling framework generate entries into each sub-population. B) Our full modelling framework, eqs (6.2.21) – (6.2.32), including the PDEs, and their boundary and initial conditions. C) The generation time distribution at calendar time t, described by $g(t, a_E, a_I, a_W, a_V)$, is obtained by normalising the instantaneous kernel $K(t, a_E, a_I, a_W, a_V)$ using $R(t)$. $R(t)$ can be estimated using between-stage reproduction numbers and generation times using ages (orange) at intermediate calendar times (blue) when generation events (i.e., stage transitions) occur.

of a mosquito/human). For example, $E(t = 3, a_E = 2)$ denotes the density of humans who were infected two time units (e.g., days) before calendar time $t = 3$. Each first-order linear equation, eqs (6.2.1) – (6.2.2), posits that, because age a_C (defined as time since entry to a sub-population C) and calendar time t are measured in the same units, the rate at which a sub-population ages would equal the rate of population change with respect to time in the absence of any instantaneous death rate (i.e., the rate at which individuals leave a sub-population C). See Murray [603] Chapter 1.7 for more details.

We close the system with the following boundary conditions:

$$E(t,0) = \int_0^\infty b_{ME}(t,a_M)M(t,a_M)da_M, \tag{6.2.3}$$

$$E(0,a_E) = f_E(a_E), \tag{6.2.4}$$

$$M(t,0) = \int_0^\infty b_{EM}(t,a_E)E(t,a_E)da_E, \tag{6.2.5}$$

$$M(0,a_M) = f_M(a_M), \tag{6.2.6}$$

where $f_E(a_E)$ and $f_M(a_M)$ denote the initial densities of infected human and mosquito populations, and each $b_{CD}(t,a_C) \geq 0$ is a time- and age-dependent birth function that represents the rate at which new individuals in sub-population D are generated at calendar time t by an individual of age a_C in sub-population C. We can solve this system using the method of characteristics – a method that reduces our first-order linear PDE to a family of ODEs.

Deriving a renewal equation and the time-varying reproduction number

To solve the system of PDEs, eqs (6.2.1) – (6.2.6), we use i) the characteristic form of the infected mosquito population, ii) the boundary condition on the mosquito population, eq. (6.2.5), and iii) and the characteristic form of the infected human population to derive the following renewal equation (see Appendix E.2):

$$E(t,0) = \int_0^t \int_0^{t-a_M} b_{ME}(t,a_M)e^{-\mu a_M} b_{EM}(t-a_M,a_E)E(t-a_M-a_E,0)da_E da_M, \tag{6.2.7}$$

where we have ignored initial conditions f_E and f_M (Eqs (6.2.4) and (6.2.6)) and assumed that the epidemic is well-established, allowing us to focus on longer-term dynamics (such that $t \gg a_M$ and $t \gg a_E$).

Then, we seek a solution of the form: $E(t,a_E) = p(a_E)e^{\gamma t}$. This solution states that the age distribution of infected humans is altered by a factor which grows or decays with time depending on whether $\gamma > 0$ or $\gamma < 0$, respectively. Substituting the expression into eq. (6.2.1), we deduce that $p(a_E) = p(0)e^{-\gamma a_E}$. From eq. (6.2.7), we have

$$\phi(\gamma) := \int_0^t \int_0^{t-a_M} b_{ME}(t,a_M)e^{-\mu a_M} b_{EM}(t-a_M,a_E)e^{-\gamma(a_E+a_M)} da_E da_M. \tag{6.2.8}$$

This expression is defined for each calendar time t. From eq. (6.2.8), we define an instantaneous growth rate $r_t := \{\gamma \text{ such that } \phi(\gamma) = 1\}$, which is unique as $\phi(\gamma)$ is a monotonically decreasing function of γ. This instantaneous growth rate for human infections provides a measure of the long-run behaviour of human infections, should the conditions affecting human transmission at time t remain the same. Our model-based definition of r_t does not account for transient behaviour (which is likely important in stochastic, early outbreak stages[104]) and should not be confused with empirical approaches (e.g.,[604]) that estimate an instantaneous growth rate r_t from case/prevalence data, often using a smoothing window approach[134]. As $\phi(\gamma)$ is a monotonically decreasing function of γ, $\phi(0)$ is the critical threshold for epidemic growth. This is because $r_t > 0$ when $\phi(0) > 1$ and $r_t < 0$ when $\phi(0) < 1$ (See Figure 6.2 A and B). So, we define $R(t)$ as:

$$R(t) := \phi(0) = \int_0^t \int_0^{t-a_M} b_{ME}(t,a_M)e^{-\mu a_M} b_{EM}(t-a_M,a_E) da_E da_M. \tag{6.2.9}$$

$R(t)$ defines a time-varying human-to-human reproduction number because our renewal equation for the newly infected human population (eq. (6.2.7)) has only a single birth state (for new human infections), and so the corresponding next-generation operator[116], describing the number of new human infections produced by previously infected individuals, is:

$$\mathbb{K} := \int_0^t \int_0^{t-a_M} b_{ME}(t,a_M)e^{-\mu a_M} b_{EM}(t-a_M,a_E) da_E da_M.$$

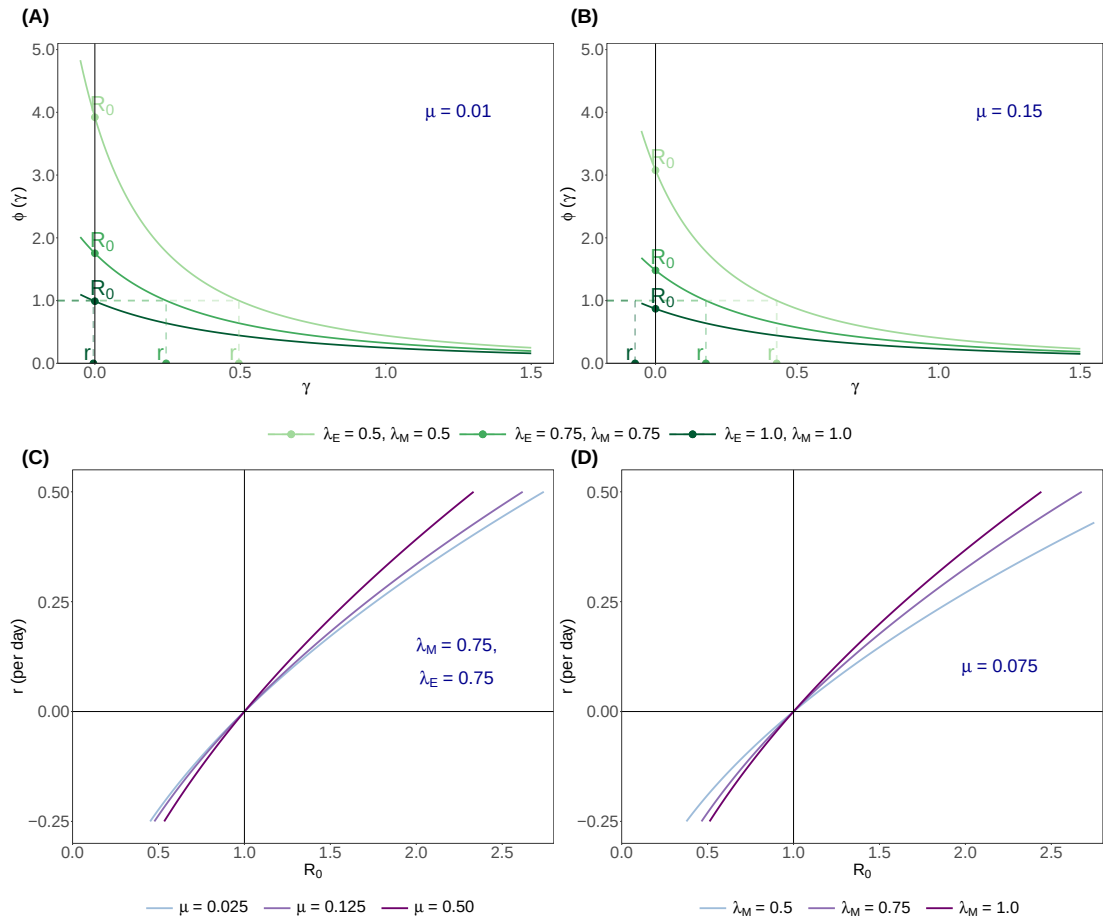


Figure 6.2: R_0 **and** r **under different mosquito-human traits** for the modelling framework described by eqs (6.2.1)-(6.2.5). We assume $b_{EM}(t, a_E) = \exp(-\lambda_E a_E)$ and $b_{ME}(t, a_M) = \exp(-\lambda_M a_M)$. These assume that the birth processes decline only with time since infection. Panels **(A)** and **(B)**: We visualise how R_0 and r are derived for varying infection rates (λ_M and λ_E for the production of infected humans and mosquitoes respectively) and mosquito death rates (μ). $r = \{\gamma$ such that $\phi(\gamma) = 1\}$ and $R_0 = \phi(0)$. Panels **(C)** and **(D)**: The relationship between r and R_0, eq. (E.2.25) in Appendix E.2.4, is traced out for the modelling framework under three different mosquito death rates μ (left) and three fixed different mosquito infection rates λ_M (right). These figures would be visualisations of the relationship between $R(t)$ and r_t, at each calendar time t, if we had not assumed time-independent birth processes $b_{EM}(t, a_E)$ and $b_{ME}(t, a_M)$.

The time-varying reproduction number $R(t)$ is the largest (and only) absolute eigenvalue of $\mathbb{K}$. So, eq. (6.2.9) defines the instantaneous human-to-human reproduction number $R(t)$ as, relying only on information up to time t, it is the average number of secondary human infections that an infected human would generate over their infected lifetime if the current conditions remained the same. Here, when discussing current conditions, these are the conditions that currently affect infection of humans – conditions which are determined by the historical conditions experienced for the previous stages of the transmission cycle.

A closer look at R_0 and r

If the birth processes $b_{ME}(t, a_M)$ and $b_{EM}(t, a_E)$ do not change over calendar time, and we take the limit as $t \to \infty$, then eq. (6.2.9) defines a basic reproduction number R_0 with a corresponding intrinsic growth rate r. We can use simple exponential forms for birth processes, $b_{ME}(t, a_M) = \exp(-\lambda_M a_M)$ and $b_{EM}(t, a_E) = \exp(-\lambda_E a_E)$, to obtain analytical results. We illustrate these derivations in Figure 6.2 for R_0 and r (see also Appendix E.2.4 for further details). The figure shows the familiar threshold relationship between R_0 and r i.e., $R_0 > 1$ when $r > 0$. Also, faster declines in infection rates

(from time since infection) and higher mosquito death rates both increase the exponential growth rate r for a fixed $R_0 > 1$. Intuitively, this is because if, for example, mosquitoes have shorter lives, yet bequeath the same number of infections over their infected lifetime, then the epidemic must be growing faster. The figure also demonstrates that interventions which reduce R_0 by a fixed amount cause smaller changes in r as R_0 increases.

Instantaneous kernels and generation time distributions

$R(t)$ is often written in terms of i) the instantaneous kernel or ii) the generation time distribution and the instantaneous growth rate. To obtain this form, we write eq. (6.2.9) as

$$\begin{aligned} R(t) &= \int_0^t \int_0^{t-a_M} b_{ME}(t, a_M) e^{-\mu a_M} b_{EM}(t - a_M, a_E) da_E da_M, \\ &= \int_0^t \int_0^{t-a_M} K(t, a_M, a_E) da_E da_M, \end{aligned} \tag{6.2.10}$$

which suggests defining a 2-dimensional probability density, at calendar time t, for the generation time $g(t, a_M, a_E) := K(t, a_M, a_E)/R(t)$. This distribution controls the times between the primary and secondary human infections. As required for a generation time distribution, it is the normalised version of the instantaneous kernel $K(t, a_M, a_E)$ [133,145,605]. The kernel defines the average rate at which secondary human infections are generated at time t by i) a mosquito infected a_M time units ago and ii) a human that was of infection age a_E at time $t - a_M$. So, $g(t, a_M, a_E)$ defines the relative contributions to secondary human infections at time t of a mosquito of infection age a_M at time t that was itself infected by a human of age a_E at time $t - a_M$. This is similar to existing interpretations for directly transmitted infectious diseases, as the expected relative contribution to the force of infection by a human infected a $\tau = a_E + a_M$ time units ago [145,606].

As in the setting of directly transmitted infectious diseases [145], both $g(t, a_M, a_E)$ and $R(t)$ are (counterfactual) instantaneous measures of relative contributions to human infections and of time-varying transmissibility respectively; these are counterfactual because they are defined using historical conditions under assumptions of conditions affecting human transmission remaining the same as at time t.

Using eqs (6.2.8) and (6.2.10), we have:

$$\frac{1}{R(t)} = \int_0^t \int_0^{t-a_M} g(t, a_M, a_E) e^{-r_t (a_E + a_M)} da_E da_M, \tag{6.2.11}$$

which shows how $g(t, a_M, a_E)$ controls the relationship between r_t and $R(t)$ (Figure 6.2 C and D).

Our renewal equation for the newly infected human population can now be written in terms of the generation time distribution and the instantaneous reproduction number:

$$E(t, 0) = \int_0^t \int_0^{t-a_M} K(t, a_M, a_E) E(t - a_M - a_E, 0) da_E da_M, \tag{6.2.12}$$

$$= R(t) \int_0^t \int_0^{t-a_M} g(t, a_M, a_E) E(t - a_M - a_E, 0) da_E da_M. \tag{6.2.13}$$

The renewal equation features $E(t - a_M - a_E, 0)$, the humans previously infected at time $t - a_M - a_E$. The corresponding generation time distribution is described by $g(t, a_M, a_E)$ and is calculated using intermediate calendar times when generation events occur in the transmission cycle (see eq. (6.2.20) below). We define a generation event to occur when a new individual is generated (or born) in a sub-population.

The key to our generation time distribution is that it depends on i) the stages of the transmission cycle and ii) the calendar times at which they begin – see Figure 6.1 C for a visualisation.

Between-stage reproduction numbers

We can also define a reproduction number between stages, which gives the average number of infected humans generated per infected mosquito:

$$\begin{aligned} R_{ME}(t) &:= \int_0^t b_{ME}(t, a_M) e^{-\mu a_M} da_M, \\ \beta_{ME}(t, a_M) &:= \frac{b_{ME}(t, a_M) e^{-\mu a_M}}{R_{ME}(t)}, \end{aligned} \tag{6.2.14}$$

where $1 = \int_0^t \beta_{ME}(t, a_M) da_M$ and $\beta_{ME}(t, a_M)$ is the probability density, at calendar time t, of the generation time from infected mosquitoes to infected humans. $\beta_{ME}(t, a_M)$ is defined by normalising the instantaneous kernel between stages (i.e., the average rate at which human infections are generated at time t by a mosquito infected a_M time units ago). So, $\beta_{ME}(t, a_M)$ defines the relative contribution of mosquitoes of age a_M to human infections at time t.

We also produce a reproduction number for the human-to-mosquito component of the dynamics:

$$\begin{aligned} R_{EM}(t - a_M) &:= \int_0^{t-a_M} b_{EM}(t - a_M, a_E) da_E, \\ \beta_{EM}(t - a_M, a_E) &:= \frac{b_{EM}(t - a_M, a_E)}{R_{EM}(t - a_M)}, \end{aligned} \tag{6.2.15}$$

where $1 = \int_0^{t-a_M} \beta_{EM}(t - a_M, a_E) da_E$ and $\beta_{EM}(t - a_M, a_E)$ is the probability density, at calendar time $t - a_M$, of the generation time from infected humans to infected mosquitoes. So, $\beta_{EM}(t - a_M, a_E)$ describes the relative contributions of infected humans (to the generation of newly infected mosquitoes) that were of age a_E at time $t - a_M$.

Eqs (6.2.14)-(6.2.15) collectively result in a modification of eq. (6.2.7) to

$$E(t, 0) = R_{ME}(t) \int_0^t \beta_{ME}(t, a_M) R_{EM}(t - a_M) \int_0^{t-a_M} \beta_{EM}(t - a_M, a_E) E(t - a_M - a_E, 0) da_E da_M, \tag{6.2.16}$$

and also of eq. (6.2.9) to

$$\begin{aligned} R(t) &= R_{ME}(t) \int_0^t \beta_{ME}(t, a_M) R_{EM}(t - a_M) da_M \\ &= R_{ME}(t) \mathbb{E}_{ME}[R_{EM}(t - a_M)], \end{aligned} \tag{6.2.17}$$

where $\mathbb{E}_{ME}[X(t - a_M)] := \int_0^{t-a_M} X(t - a_M) \beta_{ME}(t, a_M) da_M$. This defines the reproduction number as expectations across the between-stage reproduction numbers. Eq. (6.2.17) is similar to a geometric average of the between-stage reproduction numbers for an entire system (e.g.,[119,122,607]), and here we define a human-to-human reproduction number $R(t)$ which requires that a new human infection involves a two-stage process.

Relation to existing renewal equation frameworks

Our integral equation, eq. (6.2.13), can be related to the ubiquitous renewal equation frameworks for directly transmitted infectious diseases (see Appendix E.2.3).

Since $t - (a_M + a_E)$ occurs in our renewal equation, eq. (6.2.13), we can define a new variable, $\tau = a_E + a_M$, measuring the overall delay between primary and secondary human infections. By a change of variables and an integration step, we produce a one-dimensional renewal equation:

$$E(t, 0) = R(t) \int_0^t w(t, \tau) E(t - \tau, 0) d\tau, \tag{6.2.18}$$

where $w(t, \tau) := \int_0^\tau g(t, \tau - a_E, a_E) da_E$.

If we assume the generation time distribution is constant over calendar time, (i.e., $g(t, a_M, a_E) = g(a_M, a_E), \forall t$), we recover the familiar renewal equation [137,139,148,149], as in eq. (6.1.1):

$$E(t,0) = R(t) \int_0^t w(\tau) E(t-\tau, 0) d\tau. \tag{6.2.19}$$

Assuming a constant generation time distribution over calendar time means that the ratio

$$g(t, a_M, a_E) = \frac{K(t, a_M, a_E)}{R(t)} = \frac{b_{ME}(t, a_M) e^{-\mu a_M} b_{EM}(t - a_M, a_E)}{\int_0^t \int_0^{t-a_M} b_{ME}(t, a_M) e^{-\mu a_M} b_{EM}(t - a_M, a_E) da_E da_M}, \tag{6.2.20}$$

is fixed over calendar time (t). So, this typically implicit modelling assumption means that the relative expected contribution to secondary human infections of each pair of ages (i.e., times-in-states (a_M, a_E)) is constant over the course of an outbreak. This means $K(t, a_M, a_E)$ scales linearly over calendar time t with the expected overall number of new infections generated (across all age combinations of sub-populations) per initially infected human, $R(t)$.

The assumption may be unrealistic for dengue or any vector-borne disease with a multi-stage transmission cycle. For example, higher temperatures may reduce the extrinsic incubation period of dengue [80], and/or changes in the host and vector population sizes (which may themselves have temperature and precipitation dependencies) could alter the distribution of times between generation events.

Eq. (6.2.19) formalises the relationship between our framework and existing renewal equations (used in e.g., [137,139,148,149]). The existing renewal equations are a special case of our general framework with an additional assumption of a time-invariant generation time distribution.

6.2.2 Exposure-and-infection-age-structured populations of human and mosquitoes

Modelling framework

We now extend the above framework to a transmission cycle with four sub-populations which includes exposed and infectious sub-populations in the host and vector populations. This structure accounts for the intrinsic and extrinsic latent periods (i.e., the time taken from mosquito bite to the host or mosquito respectively becoming infectious). The system, consisting of densities of an exposed (E) and infectious (I) human population, and an exposed (W) and infectious (V) mosquito population, is subject to the following conservation equations (e.g., Chapter 1.7 in Murray [603]):

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a_E} = -b_{EI}(t, a_E) E(t, a_E), \tag{6.2.21}$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a_I} = 0, \tag{6.2.22}$$

$$\frac{\partial W}{\partial t} + \frac{\partial W}{\partial a_W} = -b_{WV}(t, a_W) W(t, a_W) - \mu W(t, a_W), \tag{6.2.23}$$

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial a_V} = -\mu V(t, a_V), \tag{6.2.24}$$

where again each $b_{CD}(t, a_C) \geq 0$ is a time- and age-dependent birth function that represents the rate at which new individuals in sub-population D are generated at calendar time t by an individual of age a_C in sub-population C. $\mu \geq 0$ is the mosquito death rate for both exposed and infectious mosquitoes, which we assume is age- and infection-status-independent [608].

We close the system with the following initial and boundary conditions:

$$E(t,0) = \int_0^\infty b_{VE}\left(t,a_V\right) V\left(t,a_V\right) da_V, \tag{6.2.25}$$

$$E\left(0,a_E\right) = f_E\left(a_E\right), \tag{6.2.26}$$

$$I(t,0) = \int_0^\infty b_{EI}\left(t,a_E\right) E\left(t,a_E\right) da_E, \tag{6.2.27}$$

$$I\left(0,a_I\right) = f_I\left(a_I\right), \tag{6.2.28}$$

$$W(t,0) = \int_0^\infty b_{IW}\left(t,a_I\right) I\left(t,a_I\right) da_I, \tag{6.2.29}$$

$$W\left(0,a_W\right) = f_W\left(a_W\right), \tag{6.2.30}$$

$$V(t,0) = \int_0^\infty b_{WV}\left(t,a_W\right) W\left(t,a_W\right) da_W, \tag{6.2.31}$$

$$V\left(0,a_V\right) = f_V\left(a_V\right). \tag{6.2.32}$$

where $f_E(a_E)$, $f_I(a_I)$, $f_W(a_W)$, and $f_V(a_V)$ are the initial densities for the exposed human, infectious human, exposed mosquito, and infectious mosquito populations respectively. Note that $b_{\mathrm{VE}}(t,a_V)$ and $b_{\mathrm{IW}}(t,a_I)$ represent the processes by which new infections occur in humans and mosquitoes respectively, while $b_{\mathrm{EI}}(t,a_I)$ and $b_{\mathrm{WV}}(t,a_V)$ represent the processes by which infected humans and mosquitoes become infectious. We consider possible functional forms for these birth processes in eq. (6.3.1).

An intermediate system with three sub-populations is provided in Appendix E.3 and additional derivation details for this four-compartment framework are provided in Appendix E.4.

Deriving a renewal equation and the time-varying reproduction number

We can derive the following renewal equation for newly infected humans:

$$\begin{aligned} E(t,0) = \int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} & b_{VE}\left(t,a_V\right) b_{WV}\left(t-a_V,a_W\right) b_{IW}\left(t-a_V-a_W,a_I\right) \\ & \cdot b_{EI}\left(t-a_V-a_W-a_I,a_E\right) e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} \\ & \cdot E(t-a_V-a_W-a_I-a_E,0) e^{-\int_0^{a_E} b_{EI}(t-a_V-a_W-a_I-a_E+s,s)ds} da_E da_I da_W da_V, \end{aligned} \tag{6.2.33}$$

where we have again assumed long-term epidemic dynamics (such that $t \gg a_V$, $t \gg a_W$, and $t \gg a_I$, $t \gg a_E$). We again try a separable solution (to derive analytical results) of the form $E(t,a_E) = p(a_E)e^{\gamma t}$. To have a separable form for eq. (6.2.33), we must assume that $b_{EI}\left(t,a_E\right) = b_{EI}\left(a_E\right) \forall t$. This amounts to assuming that the rate at which an exposed individual becomes infectious is independent of calendar time (t) and depends only on time since exposure (a_E). Then, we can derive a formula for the time-varying human-to-human reproduction number:

$$\begin{aligned} R(t) = \int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} & b_{VE}\left(t,a_V\right) b_{WV}\left(t-a_V,a_W\right) b_{IW}\left(t-a_V-a_W,a_I\right) \\ & \cdot b_{EI}\left(a_E\right) e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} e^{-\int_0^{a_E} b_{EI}(s)ds} \\ & da_E da_I da_W da_V \end{aligned} \tag{6.2.34}$$

$$= \int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} K(t,a_E,a_I,a_W,a_V) da_E da_I da_W da_V,$$

which suggests the following definition of a four-dimensional probability density

$$g(t,a_E,a_I,a_W,a_V) := \frac{K(t,a_E,a_I,a_W,a_V)}{R(t)}, \tag{6.2.35}$$

for the generation time at calendar time t, derived by normalising the instantaneous kernel.

Again, the time-varying reproduction number $R(t)$ is an instantaneous reproduction number for the infected human population, and defines the expected number of secondary human infections produced at calendar time t by an infected human, if the conditions that affect human infections at time t were to remain the same. The instantaneous kernel $K(t, a_E, a_I, a_W, a_V)$ is the expected rate at which secondary human infections are generated at calendar time t across individuals of stage-specific ages (a_E, a_I, a_W, a_V); these ages are the times-since-state-entry at the calendar times when each sub-population plays their role in the transmission cycle.

At each calendar time t, the generation time distribution describes the relative contributions to the current number of new human infections generated by previously exposed humans, previously infectious humans, previously exposed mosquitoes, and currently infectious mosquitoes of specific ages (i.e., the times spent in compartments until each population plays their role in generation events – see Figure 6.1 C).

We can then rewrite eq. (6.2.33) to produce a more familiar renewal equation:

$$E(t,0) = R(t) \int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} g(t, a_E, a_I, a_W, a_V) \cdot E(t - a_V - a_W - a_I - a_E, 0) da_E da_I da_W da_V. \tag{6.2.36}$$

Between-stage reproduction numbers

As before, we can define reproduction numbers and generation time distributions between stages (see Appendix E.4.3). This allows us to rewrite eq. (6.2.34) as an expectation of reproduction numbers across stages of the transmission cycle:

$$R(t) = R_{VE}(t)\mathbb{E}_{VE}\left[R_{WV}(t - a_V)\mathbb{E}_{WV}\left[R_{IW}(t - a_V - a_W)\mathbb{E}_{IW}\left[R_{EI}\left(t - a_V - a_W - a_I\right)\right]\right]\right], \tag{6.2.37}$$

where each $\mathbb{E}_{AB}[R_{CD}(t - a_A)] := \int_0^{t-a_A} R_{CD}(t - a_A)\beta_{AB}(t, a_A)da_A$ and each $R_{CD}(t)$ defines the average number of individuals in sub-population D produced by an individual in sub-population C if conditions were to remain the same as at time t. The between-stage reproduction numbers are evaluated with upper limits determined by the calendar times at which the subsequent stage of the transmission cycle begins. The expectations are taken with respect to the generation time distribution (β_{AB}) between stages of the next stage of the transmission cycle. The generation time distributions between stages are evaluated using ages of the sub-populations at the calendar times at which onward generation events in the transmission cycle occur (see Figure 6.1 C: a_V at time t, a_W at time $t - a_V$, and a_I at time $t - a_V - a_W$).

Application to simulated and real-world data

We now consider a numerical application, where our aim was to investigate the importance of tracking the calendar times of the transmission cycle when estimating the generation time distribution. For simplicity, this example is not a full exposition of our mechanistic framework (eqs (6.2.21) – (6.2.32)); it only uses the assumptions governing how stage durations depend on temperature outlined by Siraj et al.[156] (see Appendix E.4.4), and does not consider estimating $R(t)$ (see Discussion). All code and data used in these applications are contained in a GitHub repository[610].

We used daily mean temperature data from Ang Mo Kio (a town in Singapore), Foz do Iguaçu (a city in Brazil), and São João do Tauape (a neighbourhood in the city of Fortaleza, Brazil), as well as simulated daily temperature data[282,609]. Locations were selected due to known circulation of dengue viruses. We investigated differences in generation time distributions when using a stage-specific approach versus an existing one (see Siraj et al.[156]) which ignores the calendar times of the stages of the transmission cycle. For both approaches, we estimated the generation time distribution assuming the four-stage transmission cycle (Figure 6.1 A) and Monte Carlo sampling (see Appendix E.4.4). Our Monte Carlo procedure for both approaches provides samples of $\tau = a_E + a_I + a_W + a_V$, the time between primary and secondary human

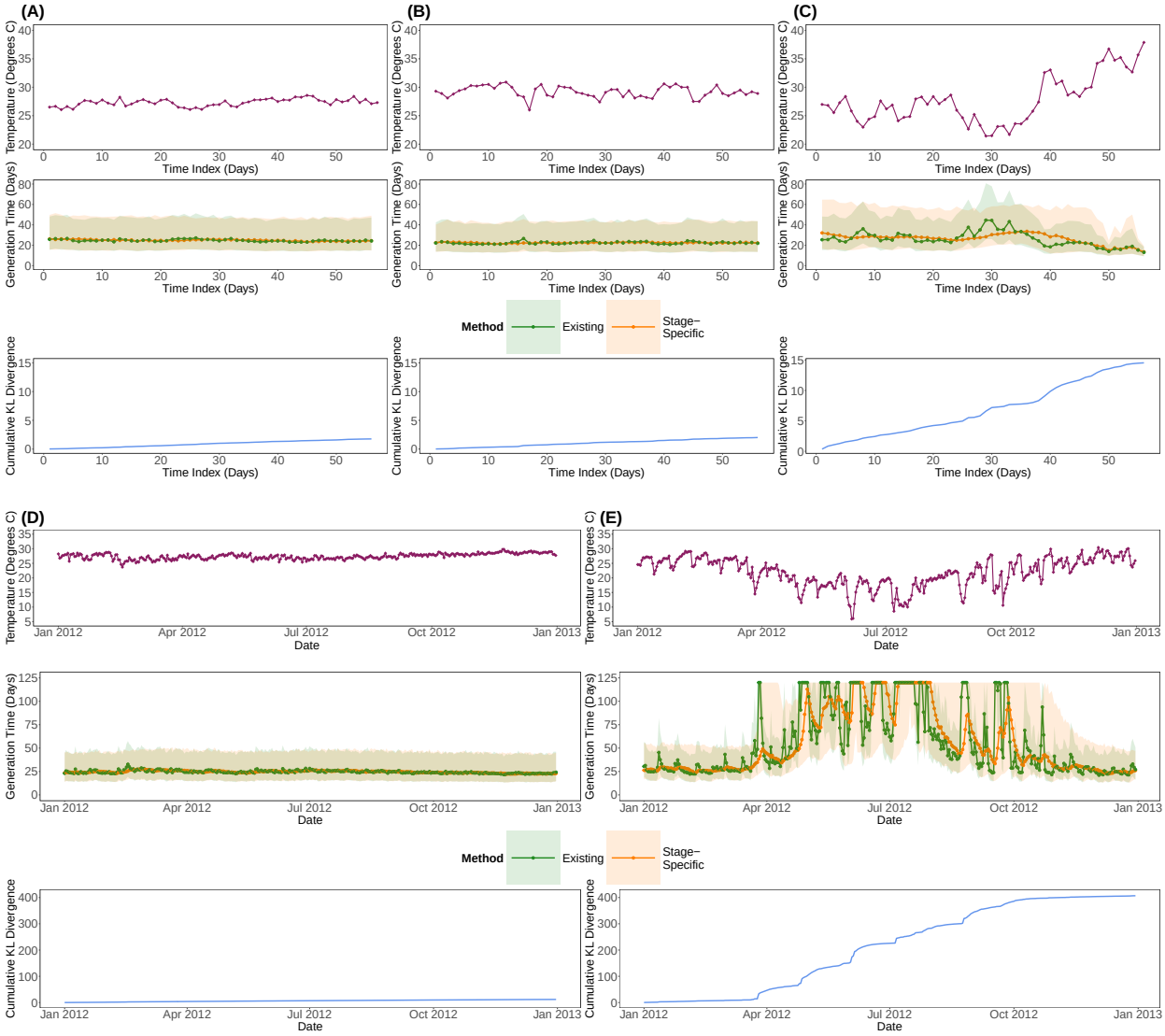


Figure 6.3: **Generation time (GT) distributions under temperature-dependent conditions:** For the four-stage transmission cycle of dengue, we illustrate GT distributions at different calendar times. In these toy examples, the distributions were estimated using Monte Carlo sampling (see Appendix E.4.4). We assumed temperature-dependent GT distributions and used daily temperature data and either a stage-specific (orange) or an existing (green) approach. The filled circles indicate the medians of the 2000 Monte Carlo samples, and the shaded regions indicate the central 95% intervals of the samples. For the *stage-specific* approach, the GT distribution was based on temperatures observed for calendar times when the estimated generation events occurred (i.e., depending on current and historical temperatures). The *existing* approach ignores the calendar times at which transitions between stages of the transmission cycle occur, and the GT distribution is changed every day based on the temperature observed that day. For both approaches, we assumed the same temperature effects on times between transmission cycle stages (which were outlined by Siraj et al. [156], see Appendix E.4.4). GT distributions were estimated using daily mean temperature (purple) data for (A) Ang Mo Kio, Singapore (December 2023 to June 2024), B) São João do Tauape, Brazil (November 2011 to May 2012), (C) a simulated environment (using a Gaussian random walk with noise $\epsilon_t \sim N(0, 4)$, and maximum and minima of 20 °C and 45 °C respectively), (D) São João do Tauape, Brazil (January 2012 to December 2012), and (E) Foz do Iguaçu, Brazil (January 2012 to December 2012) [282,609]. The upper plots, (A)-(C), show GT distributions over 56 days (8 weeks), while the lower plots, (D)-(E), show GT distributions over one year. We used different y-axes for the plots in panels (A)-(C) vs (D)-(E) for legibility. For panel (E), we bounded the maximum GT delay (y-axis) at 120 days, consistent with the approach taken by Siraj et al. [156]. For each of (A)–(E), we measured the difference between the existing and stage-specific approach using Kullback–Leibler (KL) divergence, and plotted the cumulative KL divergence over time.

infections (which follows the distribution with density $w(t, \tau)$). To produce a sample of τ, we sample a_E, a_I, a_W, and a_V from their assumed between-stage distributions.

For i) the existing approach which ignores the calendar times at which transitions between stages of the transmission cycle occur, the generation time distribution is changed every day based on the temperature observed that day. To obtain samples from $w(t,\tau)$, we sampled $a_E^i \sim \beta_{EI}(t, a_E)$, $a_I^i \sim \beta_{IW}(t, a_W)$, $a_W^i \sim \beta_{WV}(t, a_W)$, and $a_V^i \sim \beta_{VE}(t, a_V)$, based on the temperature observed at the same calendar time t. This Monte Carlo sampling procedure gives a single sample $\tau^i = a_E^i + a_I^i + a_W^i + a_V^i$ from $w(t,\tau)$, and we repeated this procedure to obtain 2000 samples from $w(t,\tau)$.

For ii) our stage-specific approach, the generation time distribution was based on temperatures observed for calendar times when the estimated generation events occurred. Both approaches have the same assumptions for temperature effects on stage durations, but our stage-specific approach uses between-stage distributions based on current *and* historical temperatures. For this stage-specific approach, we used the sampling distributions (between stages) outlined by Siraj et al.[156] in the following way: we sampled $a_V^i \sim \beta_{VE}(t, a_V)$, $a_W^i \sim \beta_{WV}(t - a_V^i, a_W)$, $a_I^i \sim \beta_{IW}(t - a_V^i - a_W^i, a_W)$, and $a_E^i \sim \beta_{EI}(t - a_V^i - a_W^i - a_I^i, a_E)$. This Monte Carlo sampling procedure gives a single sample $\tau^i = a_E^i + a_I^i + a_W^i + a_V^i$ from $w(t,\tau)$. Note a sample for a_E is based on the between-stage distribution at time t, whereas samples of a_I, a_W, and a_V are based on the between-stage distributions at times before time t. We repeated this procedure to obtain 2000 samples from $w(t,\tau)$.

We quantified the difference between the existing and stage-specific approaches at each time point using the Kullback-Leibler (KL) divergence, and visualised the cumulative KL divergence to capture the total difference over time.

We investigated short time periods (eight weeks, Figure 6.3 A-C) where dengue was likely to circulate and longer time periods (an entire year, Figure 6.3 D-E) where dengue was less likely to circulate (due to winter temperatures). To illustrate the effects of a volatile environment, our simulated daily temperature data were based on a Gaussian random walk (with noise $\epsilon_t \sim N(0, 4)$, and reflecting boundaries at 20 °C and 45 °C).

For the eight-week period, in the applications to real-world temperature data (Figure 6.3 A-B, Figures SI E.1 – SI E.4), there were similar trends in the estimated generation time distributions across the two methods. However, the higher variability in the simulated temperature data (Figure 6.3 C, Figures SI E.5 – SI E.6) resulted in notable differences in the evolution of the generation time distributions, as indicated by the cumulative KL divergence over the eight-week period. For the year-long period, in São João do Tauape (Figure 6.3 D), the generation time distribution from both methods did not fluctuate wildly over calendar time since the temperature was relatively stable temperature conditions. In contrast, the volatile temperatures observed in the city of Foz do Iguaçu (Figure 6.3 E) resulted in extreme, long generation times. These results hint that at lower temperatures, transmission could be slowed by temperature-mediated effects. However, these estimates are the result of extrapolating laboratory results directly into the field, which likely overstates the impact of temperature on transmission (see Discussion). Changes in temperature are also likely correlated with rainfall amounts, resulting in changes in the mosquito population dynamics which may dominate the temperature effects.

Across the five settings (Figure 6.3 A-E), the stage-specific approach acted like a smoothed average versus the existing approach as the generation time distribution was based on temperatures observed throughout the transmission cycle. These differences are most pronounced in the simulated setting (Figure 6.3 C) and Foz do Iguaçu (Figure 6.3 E) when temperatures varied more widely, and the cumulative KL divergence was substantially larger in these settings versus analogous settings with less temperature variability.

6.3 Discussion

Transmission of mosquito-borne pathogens is sensitive to environmental fluctuations, particularly to changes in temperature, and elements of their transmission cycle respond differently to these variations[80,82,88,580]. This temporal variation means that there is no fixed distribution governing the time it takes one human infection to cause (through the vector) a subsequent human infection. This creates issues for traditional renewal frameworks which assume that this so-called generation time distribution is fixed. Here, we derive from first principles a renewal equation that applies to mosquito-borne diseases, which

naturally accounts for temporal variation in each element of the transmission cycle. Time-varying reproduction numbers and epidemic growth rates emerge from our renewal equation framework, and this provides a theoretically grounded basis for estimating these important metrics of pathogen transmission rates.

Our empirical results indicate that, under the assumption of previously estimated temperature dependencies (for times between stages of the transmission cycle)[80,156], the generation time distribution is likely to be transient, changing during a mosquito-borne disease outbreak in response to varying environmental conditions. Future empirical work may look at embedding other environmental, biological, and demographic conditions (beyond temperature) into the time-varying components of the birth processes which control the generation time distribution (see below, eq. (6.3.1)). However, our results indicate that this distribution depends on both current and historical conditions. This dependence on history is because for transmission to occur, a full sequence of stages of the transmission cycle must be completed, and each of these stages could be exposed to different conditions. Our generation time distribution at a given time then represents the uncertainty in the typical time between parent and daughter infections if conditions were as they were historically. The mechanistic, history-dependent nature of our generation time distribution contrasts with existing less mechanistic approaches that more heuristically allow the generation time distribution to change in response to only current conditions[156]. Our empirical results (albeit only using previously documented temperature dependencies for times between stages of transmission[80,156]) hint that our approach predicts less temporal variation in the generation time distribution than approaches based only on current conditions. This is because the dependence on historical conditions likely acts to smooth its variation over time. More generally, beyond our simplified numerical application, our theory for the generation time distribution also differs from other less mechanistic approaches[375,376,377]. These approaches either assume time-invariant generation time distributions or do not directly normalise the relative contributions of different sub-population ages (i.e., the instantaneous kernel), and instead can only respond to observed temperature values – ignoring other sources of temporal variation for elements of the transmission cycle.

Ours and other approaches to modelling mosquito-borne pathogen transmission depend on assumptions about how elements of the pathogen and mosquito lifecycle respond to environmental conditions, and these are largely drawn from laboratory studies. Many such studies have shown that variations in temperature substantially affect the time spent in particular lifecycle stages. For malaria, Zika, dengue, and a range of arboviruses, the time taken for an infected mosquito to become infectious is hastened by higher environmental temperatures in the laboratory. Mosquitoes also die more rapidly at elevated laboratory temperatures[80,83,217]. These laboratory studies underpin a wide range of modelling studies that extrapolate how mosquito-borne pathogen transmission may respond to changes in temperature (e.g.,[89,156]). However, more work is required to understand whether these results from laboratory studies hold in the wild: mosquitoes, like all insects, are cold-blooded and rely on their environment to thermoregulate, meaning they adjust their behaviour in response to changes in environmental conditions. Since the laboratory mosquitoes are limited in their environment, they are likely limited in their ability to respond to temperature changes. Laboratory mosquitoes are often highly inbred which may also make them more fragile to temperature changes. This means that laboratory studies likely overstate the impacts of temperature on disease transmission, and field or semi-field studies are crucial.

Despite recent advances in understanding temperature-dependent transmission[82,611,612], we still do not have a comprehensive understanding of whether results from laboratory studies will transfer to the field. This is essential as the incidence and distribution of many mosquito-borne diseases such as dengue continues to rise in the climate crisis, and model-based projections under different climate change scenarios are often underpinned by these estimated temperature dependencies[83,154]. Understanding how mosquitoes respond (and the shape of the generation time distribution) to increasing future climatic variability will be important for future projections of dengue burden and could be achieved by embedding these projections in our model. Despite our simplifying numerical application, there are many environmental and biological factors other than temperature that may affect the transmission of pathogens. For example, the rate at

which mosquitoes become infected and infect humans depends on demography – the population sizes of mosquitoes and susceptible humans (see eq. (6.3.1)). Mosquito population dynamics are driven by the availability of water sources due to the dependence of the initial mosquito life stages on these. So, rainfall also affects the rate of pathogen transmission, although this aspect is typically ignored in existing renewal-type models of mosquito-borne infection transmission. Our model considers pathogen dynamics and only the dynamics of infected mosquitoes, and we could straightforwardly link it to the overall mosquito population dynamics by allowing the birth rate of infections in either humans or mosquitoes to depend on these. Modelling how mosquito population dynamics depend on both temperature and rainfall would then allow forward projections of pathogen transmission rates to be made.

Incorporating these dependencies (beyond temperature effects) and inferring $R(t)$ (e.g., alongside case data and/or densities of human and mosquito populations) will depend on parameterising the time-varying processes of eqs (6.2.21) – (6.2.32), which control both the generation time distribution and $R(t)$. These are key to extending our numerical application beyond estimating the generation time distribution. As an example, we could have:

$$\begin{aligned}
b_{EI}\,(t, a_E) &= \lambda^E_{\text{Latent}}(a_E) \\
b_{IW}\,(t, a_V) &= \frac{b(t)}{N^H(t)} \cdot S^M(t) \cdot p_{WI} = \frac{b(t)}{N^H(t)} \cdot \left(N^M(t) - \int_0^\infty W(t, a_W) da_W - \int_0^\infty V(t, a_V) da_V \right) \cdot p_{WI} \\
b_{WV}\,(t, a_W) &= \lambda^W_{\text{Latent}}(a_W) \\
b_{VE}\,(t, a_V) &= b(t) \cdot \frac{S^H(t)}{N^H(t)} \cdot p_{EV} = b(t) \left(N^H(t) - \int_0^t E(t, a_E) da_E - \int_0^t I(t, a_I) da_I \right) / N^H(t) \cdot p_{EV}
\end{aligned} \tag{6.3.1}$$

where $\lambda^E_{\text{Latent}}(a_E)$ and $\lambda^W_{\text{Latent}}(a_W)$ denote the rates at which latent periods end for infected humans and mosquitoes respectively, $N^H(t)$ and $N^M(t)$ denote the total human and mosquito population counts, $S^H(t)$ and $S^M(t)$ denote the susceptible human and mosquito population counts, p_{WI} and p_{EV} denote the probability of mosquito and human infection upon infectious bite respectively, and $b(t)$ denotes the rate at which a mosquito bites a human. Eq. (6.3.1) relies on counts for mosquito and human sub-populations, highlighting the usefulness of surveillance data and/or estimates for both host and vector densities. Elements of Eq. (6.3.1) also show how our applications will rely on an understanding of how mosquito, human, and viral traits respond in settings beyond the laboratory, which would involve a combination of laboratory and field/semi-field data.

Eq. (6.3.1) assumes that the infection processes $b_{VE}(t, a_V)$ and $b_{IW}(t, a_I)$ describe density-dependent transmission. Similarly, we can see how elements of the Ross-Macdonald theory and its extensions [104,121] could be incorporated in our framework, primarily via the processes of eq. (6.3.1). While the set of Ross-Macdonald models often vary in their structure, we could, for example, embed many core assumptions [121] such as a fixed human population count (N^H), latent periods (λ_{Latent}), and biting rates (b), while mosquito mortality could remain age-independent in our framework (as in eqs (6.2.21) – (6.2.32)).

Our framework is based on a deterministic model of pathogen dynamics, which models the density of each life stage by continuous variables. Because of this, it is likely most appropriate for modelling the dynamics of larger populations. In particular for small population counts, it would be useful to develop a stochastic individual-based theory of population dynamics and to explore the relationship between these results and ours, as has previously been considered for directly transmitted diseases [152,372]. These stochastic models may also offer a more natural framework for inferring their parameters (including $R(t)$) than our deterministic model.

The key advance of our research is to derive renewal equations from first principles that can be used to model host-vector disease outbreaks. Since renewal equations form the basis of methods for inferring characteristics of pathogen transmission (e.g., $R(t)$) in real time during outbreaks, we hope that this research is an initial step towards enabling public health policy advisors to track pathogen transmission more effectively during future outbreaks of a wide range of mosquito-borne diseases [613].

Chapter 7

Exact simulation of stochastic reaction processes with external variability

Cathal Mills and Ben Lambert. Exact simulation of stochastic reaction processes with external variability. June 2025

To better understand how to simulate the system of Chapter 6 and the importance of stochasticity, we needed to develop new simulation theory. In this chapter, we present a manuscript that introduces new and widely applicable theory to exactly simulate stochastic reaction processes that experience external variability and hence time-varying propensities between events[49]. The manuscript will be submitted to *The Journal of the Royal Society Interface*.

Abstract

In many branches of science, from epidemic dynamics to chemical reactions, systems are often described with a stochastic reaction processes – mathematical frameworks that capture how discrete events unfold over continuous time. The propensities are functions that control the stochastic reaction rates of these process. Conditional on knowing the current state of the system at time t, $X(t)$, the propensities can often be assumed to be fixed over time between two events. However, in many systems, there may be time variations in these propensities between events due to external influences acting on the system's evolution. Here, we derive simulation algorithms that extend the commonly used Gillespie's direct and first reaction algorithms to incorporate time-varying propensities between events. We apply the algorithms to a range of increasingly complex models. By comparing our outputs with numerically derived characteristics, we investigate the validity of our simulation algorithms. The algorithms and accompanying code of our manuscript provide a foundation for exact simulations of small stochastic reaction systems that experience external variability which influences the propensities over time.

Author contributions

- Conceptualization: **C.M**, B.L
- Methodology: **C.M**, B.L;
- Data curation: **C.M**, B.L
- Data visualisation: **C.M**, B.L

- Writing original draft: **C.M**, B.L
- Writing final draft: **C.M**, B.L

7.1 Introduction

The evolution of many dynamical systems over time is often complex and uncertain. This motivates using a *stochastic process* that allows for randomness in how the system evolves. These processes are widely used to describe phenomena such as the spread of epidemics, the evolution of stock prices, and the growth of populations and cells. Mathematically, a stochastic process is a collection of random variables over time. Stochastic processes represent a wide family of models, including branching processes, Poisson processes, and Wiener processes.[614]. Incorporating stochasticity is often essential for systems that experience randomness, especially when small population sizes or rare events cause deterministic models to become inaccurate as arguments such as the law of large numbers break down. For example, there is individual and environmental variability in the dynamics of epidemics in populations with small numbers of susceptible individuals[62]. In such scenarios, a stochastic model can produce very different dynamics compared to a corresponding deterministic model[122]. Another example is in the evolution of chemically reacting molecules over time, where stochasticity arises because quantum indeterminacy is unavoidable[615].

There are many ways to incorporate stochasticity into a model. Here, we focus on a class of stochastic processes known as continuous-time Markov processes with jumps, also known as Markov jump processes or stochastic reaction processes. These processes i) are continuous in time, ii) evolve with events describing jumps between discrete states, and iii) possess the first-order Markov property over time (i.e., the rates of the next events depend only on the current state of the system). By simulating these stochastic processes and summarising their trajectories over time, a researcher can describe how a dynamical system evolves over time. These numerical simulations are important because analytical solutions are typically available for only the simplest systems. Selecting a simulation algorithm is an important design choice. Different algorithms will have setting-specific suitability for different systems and will balance *accuracy* against computational *efficiency*, both of which will depend on properties of the system (e.g., number of states, stiff systems, etc.) and the intended use case (e.g., required accuracy, computational resources, etc.).

Building on the work of Doob[447], Daniel T. Gillespie introduced the eponymous Gillespie algorithms to exactly simulate Markov jump processes[50,51]. Gillespie's *direct* and *first reaction* algorithms are numerical schemes that are commonly used to simulate models across branches of science, such as models of chemical kinetics[616], protein dynamics[617], infectious disease spread[122], and many other complex systems. Gillespie's algorithms simulate solutions of such models; because the model is stochastic, there is a whole family of solutions that the Gillespie algorithms simulate. The algorithms are "exact" methods in the sense that they simulate the stochastic trajectory of a system. Both Gillespie algorithms produce trajectories in an iterative process for simulation, with each step determining i) the timing and ii) type of the next reaction/event. Given the current state of the system $X(t)$ at time t, the algorithms simulate the waiting time until and type of the next event given information on the constant probability $a_i(X(t))dt$ that each event i next occurs in the infinitesimal time interval $(t, t+dt)$. $a_i(X(t)$ is known as the *propensity* and is the instantaneous probability per unit of time that the event type i occurs. As a case study for this manuscript, suppose we have a simple system of chemical reactions with $M = 2$ event types representing the production/birth and degradation/death of molecules respectively. We set up this problem as:

$$X \xrightarrow{d} \Phi, \quad \Phi \xrightarrow{b} X, \tag{7.1.1}$$

where X represents a single chemical species of interest, d is the constant death rate for one molecule, b is the constant birth rate for one molecule, and Φ is a species no longer of interest. The event propensities at time t can be formed by

combining b and d with the number of molecules at time t, $X(t)$:

$$a_1(X(t)) = d \cdot X(t), \quad a_2(X(t)) = b, \tag{7.1.2}$$

where a_1 and a_2 are the event propensities for deaths and births, respectively. One step of the Gillespie algorithms uses these propensities to determine the timing and type of the next event, and the number of molecules is increased or decreased following the birth or death event.

As the Gillespie algorithms simulate one event at a time, simulations using the algorithms are often prohibitively slow for many real-world systems[615,618], particularly as the frequency of events increases, for example, in our case study as the number of species and/or counts of molecules grows. Elaborations of the Gillespie algorithms have therefore often focused on efficiency improvements. These elaborations still use the Gillespie algorithms as the theoretical basis for simulation. The next reaction algorithm of Gibson and Bruck[619] revises Gillespie's first reaction method to introduce a dependency tracking graph and a priority queue of scheduled event times. This next reaction algorithm is generally much more difficult to code[615] but improves efficiency by only requiring simulation of one uniform random number per reaction event (since it reuses previously drawn values). This makes it suitable for efficient simulation for large, complex systems[619].

Approximate methods are also commonly used to improve the efficiency of Gillespie algorithms, including τ-leaping methods which are especially important when reaction events are frequent (i.e., propensities are high)[620]. Instead of simulating one event per step of the algorithm, these approximate methods allow potentially many events to occur within a user-defined time interval of length τ before the propensity function is then updated. The accuracy and efficiency of the τ-leaping scheme is determined by τ, which can depend on the current state of the system and the rates at which events are occurring. Adaptive approaches for selecting τ based on the efficiency-accuracy trade-off are well-developed[621,622,623].

In many systems, the instantaneous probability of an event occurring per unit of time depends on both the current state of the system (e.g., the number of molecules) and external time-varying influences (e.g., environmental exposures). Then, the propensity function $a_i(t, X(t))$ for an event type i is both a function of time t *and* the current state of a system $X(t)$. In our case study, the birth and death rates for molecules become functions of time $t - b(t)$ and $d(t)$, respectively – when molecules are exposed to external variability. This means that the assumption of constant propensities between successive events breaks down. In the work that motivated this research[48], we considered the processes of humans becoming infected with dengue virus, a pathogen spread by female *Aedes* mosquitoes. The infection processes can depend on multiple time-varying influences. For example, both the mosquito and virus are sensitive to fluctuations in temperature. These temperature effects can vary over time and interact with other external environmental (e.g., precipitation), biological (e.g., changing serotypes) and human (e.g., human movement patterns) influences. As these influences vary over time between two events, the event propensities can vary between events (see Section 7.5).

When there are time-varying propensities between events, the Gillespie algorithms are not directly applicable. In Anderson[624], the authors present an alternative exact approach; using a random time change representation where event times of a system with time-varying propensities are represented using independent Poisson processes and integrated propensity functions. This approach differs from the Gillespie algorithms by *not* invoking memorylessness properties and instead uses priority queues to extend the next reaction algorithm of Gibson and Bruck[619] for time-varying propensities. This makes it a suitable exact simulation approach for large systems with time-varying propensities between events. An alternative approximate simulation method for systems with time-varying propensities[625] has added adaptive thinning[626] to the Gillespie direct algorithm[50,51], alongside an approximation based on a multivariate normal distribution. Here, we adopt a different approach as we focus on exact simulation for the original Gillespie direct and first reaction algorithms under time-varying propensities between events. We first re-derive and describe the original Gillespie algorithms for events with fixed propensities, before extending these algorithms to the events with time-varying propensities. We illustrate their usage for simulating i) simple chemical reaction systems with time-varying propensities and ii) a more elaborate example of individual-based processes in dengue transmission.

7.2 The classical Gillespie simulation algorithms for fixed propensities between events

As background for our algorithms for simulating systems with time-varying propensities between events, we retrace the original derivations of Gillespie[50,51].

7.2.1 Gillespie's direct method

Gillespie's direct method[50,51], also known as the Stochastic Simulation Algorithm (SSA), works by drawing two random numbers uniformly distributed in $(0, 1)$ to i) determine the time until the next reaction event and ii) then determine which type of event occurs.

To derive this algorithm, suppose without loss of generality that we have the same system as eq. (7.1.1) where the current state of the system $X(t)$ denotes the number of molecules at time t. We wish to determine both the time and type of the next reaction event. We first derive $\mathbb{P}_z(t+\tau, t+\tau+dt)$ – the probability that the next reaction will occur in the time interval $(t+\tau, t+\tau+dt)$ and will be a reaction event of type z (i.e., a birth or death), where τ denotes the waiting time and dt denotes an infinitesimally small time step.

Denoting the joint probability distribution of this event as $f(\tau, z)$, we have:

$$f(\tau, z)dt = \mathbb{P}(t' > t+\tau) \cdot \mathbb{P}(\text{type } z \text{ reaction occurs in } (t+\tau, t+\tau+dt)) \tag{7.2.1}$$

$$= \mathbb{P}(t' > t+\tau) \cdot a_z dt, \tag{7.2.2}$$

where $\mathbb{P}(t' > t+\tau)$ denotes the survival probability (i.e., that no reaction event occurs in the interval $[t, t+\tau)$), and a_z is the fixed propensity function for an event of type z during the infinitesimal interval $(t+\tau, t+\tau+dt)$. In our notation, we drop the dependence of a_z on the current state of the system $X(t)$ as the propensity a_z is fixed during the infinitesimal interval $(t+\tau, t+\tau+dt)$. This amounts to a_z only being a function of the current state of the system and not any external time-varying influences.

With M different events, the total propensity becomes $\lambda := \sum_{z=1}^{M} a_z$, and eq. (7.2.1) becomes:

$$f(\tau, z)dt = \exp(-\lambda \cdot \tau) \cdot a_z dt. \tag{7.2.3}$$

Using the law of total probability, we have:

$$f(\tau, z) = f(\tau) f(\tau \mid z), \tag{7.2.4}$$

where the marginal probability density for the waiting time $f(\tau)$ can be written as:

$$\begin{aligned} f(\tau) &= -S'(\tau) \\ &= -\frac{d}{d\tau}\left[\exp(-\lambda \cdot \tau)\right] \\ &= \lambda \cdot \exp(-\lambda \cdot \tau), \end{aligned} \tag{7.2.5}$$

where $S'(\tau)$ is the first derivative of the survival probability with respect to τ.

Eq. (7.2.5) makes explicit that waiting times between events are exponentially distributed with rate parameter λ. This suggests that we can use inverse transform sampling to generate waiting time τ. Inverse transform sampling is an independent sampling method that works as follows. Suppose random variable Y has cumulative distribution function (CDF)

$F(y)$. Then, inverse transform sampling uses the availability of inverse CDF F^{-1} and the fact that if $U \sim \text{Uniform}(0, 1)$, then $Y = F^{-1}(U)$ has the correct distribution F. For $\tau \sim \text{Exponential}(\lambda)$ and $r \sim \text{Uniform}(0, 1)$, we have:

$$\tau = -\frac{1}{\lambda} \log\left(\frac{1}{r}\right). \tag{7.2.6}$$

Meanwhile, using eqs (7.2.3) – (7.2.5) and Bayes' Rule, we can write:

$$\begin{aligned} f(z \mid \tau) &= \frac{f(\tau, z)}{f(\tau)} \\ &= \frac{\exp(-\lambda \cdot \tau) \cdot a_z}{\lambda \cdot \exp(-\lambda \cdot \tau)} \\ &= \frac{a_z}{\lambda}, \end{aligned} \tag{7.2.7}$$

Eq. (7.2.7) states that the probability distribution for event type z conditional on the next reaction occurring at time $t + \tau$, is equal to the fraction of the total propensity across events that is explained by the propensity for event type z.

This means we have a categorical distribution for determining the event type z given waiting time τ:

$$z \sim \text{Categorical}\left(\frac{a_1}{\lambda}, \frac{a_2}{\lambda}, \ldots \frac{a_M}{\lambda}\right). \tag{7.2.8}$$

All of this suggests the simulation approach given by Algorithm 1:

Algorithm 1 Gillespie's direct algorithm for fixed propensities

Initialise time $t = t_0$ and the current state of the system $X(t) = X_0$

1: Compute event propensities $a_j \; \forall \; j \in \{1, \ldots M\}$ and total propensity $\lambda = \sum_{j=1}^{M} a_j$
2: Draw waiting time $\tau \sim f(\tau)$ using eq. (7.2.6)
3: Draw event type $z \sim f(z \mid \tau)$ using eq. (7.2.8)
4: Set $t = t + \tau$ and change* the state of the system $X(t)$ according to event type z
5: Record $X(t)$. Return to Step 1, or end the simulation
6: **Output:** $(\mathbf{X(t)}, \mathbf{t})$

*In step 4, for our case study, we change the state of the system by increasing or decreasing the number of molecules at time t based on whether a birth or death event type was sampled.

We repeat the iterative steps 1–5 of Algorithm (1) until a pre-defined maximum time or until the system reaches a pre-specified state. This produces a trajectory of the system over time. In our case study, this is the number of molecules $X(t)$ of our studied chemical species X at different times t.

7.2.2 Gillespie's first reaction method

Gillespie's first reaction method is another exact simulation algorithm that produces the same distribution of trajectories as the direct method[615] and differs by sampling M different independent candidate waiting times for each of the M events. The minimum of these M candidate waiting times is used to determine the timing and type of next event. We will now derive the algorithm.

Each of the M candidate waiting times is exponentially distributed with probability density $f(\tau_z)$ for the waiting time of next reaction for event type z, defined as:

$$f(\tau_z) = \exp(-a_z \tau_z) \cdot a_z. \tag{7.2.9}$$

Let $\tilde{f}(\tau, z)dt$ be the probability that the first reaction method produces a reaction of type z in the interval $(t+\tau, t+\tau+dt)$. This probability can be written as:

$$\tilde{f}(\tau, z)dt = \underbrace{\mathbb{P}(\tau < \tau_z < \tau + dt)}_{(a)} \cdot \underbrace{\mathbb{P}(\tau_z > \tau, \forall j \neq z)}_{(b)}, \tag{7.2.10}$$

where, for (b), we have used the assumption that two reactions occurring in $(t+\tau, t+\tau+dt)$ is of order 0.

Considering (a) in eq. (7.2.10), we have:

$$\mathbb{P}(\tau < \tau_z < \tau + dt) = \exp(-a_z\tau_z) \cdot a_j dt, \tag{7.2.11}$$

while (b) in eq. (7.2.10) can be written as:

$$\begin{aligned} \mathbb{P}(\tau_z > \tau, \forall j \neq z) &= \prod_{\substack{j=1 \\ j\neq z}}^{M} \exp(-a_j\tau) \\ &= \exp\left(-\sum_{\substack{j=1 \\ j\neq z}}^{M} a_j\tau\right), \end{aligned} \tag{7.2.12}$$

where we have used the independence of events at a single step of the simulation routine.

We can now rewrite eq. (7.2.10) as:

$$\tilde{f}(\tau, z)dt = a_z \exp\left(-\sum_{j=1}^{M} a_j\tau\right) \tag{7.2.13}$$

$$= a_z \exp\left(-\lambda \cdot \tau\right), \tag{7.2.14}$$

which coincides with eq. (7.2.3) – the joint probability that the next reaction occurs in $(t+\tau, t+\tau+dt)$ and will be of type z is the same under the direct and first reaction methods.

For implementation, as each candidate waiting time for the M individual events is exponentially distributed (by eq. (7.2.9), $\tau_z \sim \text{Exponential}(a_z)$), this suggests using inverse transform sampling to sample each of the M candidate event waiting times:

$$\tau_z = -\frac{1}{a_z}\log(\frac{1}{r_z}), \tag{7.2.15}$$

where $r_z \sim \text{Uniform}(0, 1)$.

Gillespie's first reaction algorithm can now be implemented as:

Algorithm 2 Gillespie's first reaction algorithm for fixed propensities

Initialise time $t = t_0$ and the current state of the system $X(t) = X_0$

1: Compute event propensities $a_j, \quad \forall j \in \{1, \dots M\}$

2: Draw candidate waiting times $\tau_j \sim f(\tau_j), \quad \forall j \in \{1, \dots M\}$ using eq. (7.2.15)

3: Choose waiting time $\tau = \min\left(\{\tau_j\}\right)$ and event type $z = \arg\min_j\left(\{\tau_j\}\right)$

4: Set $t = t + \tau$ and change the state of the system $X(t)$ according to event type z

5: Record $X(t)$. Return to Step 1, or end the simulation

6: **Output:** $(\mathbf{X(t)}, \mathbf{t})$

Comparing Algorithms (1) and (2), the direct method is computationally more efficient as it involves sampling one total waiting time τ and one event type z given τ, whereas the first reaction method requires sampling of M candidate reaction times.

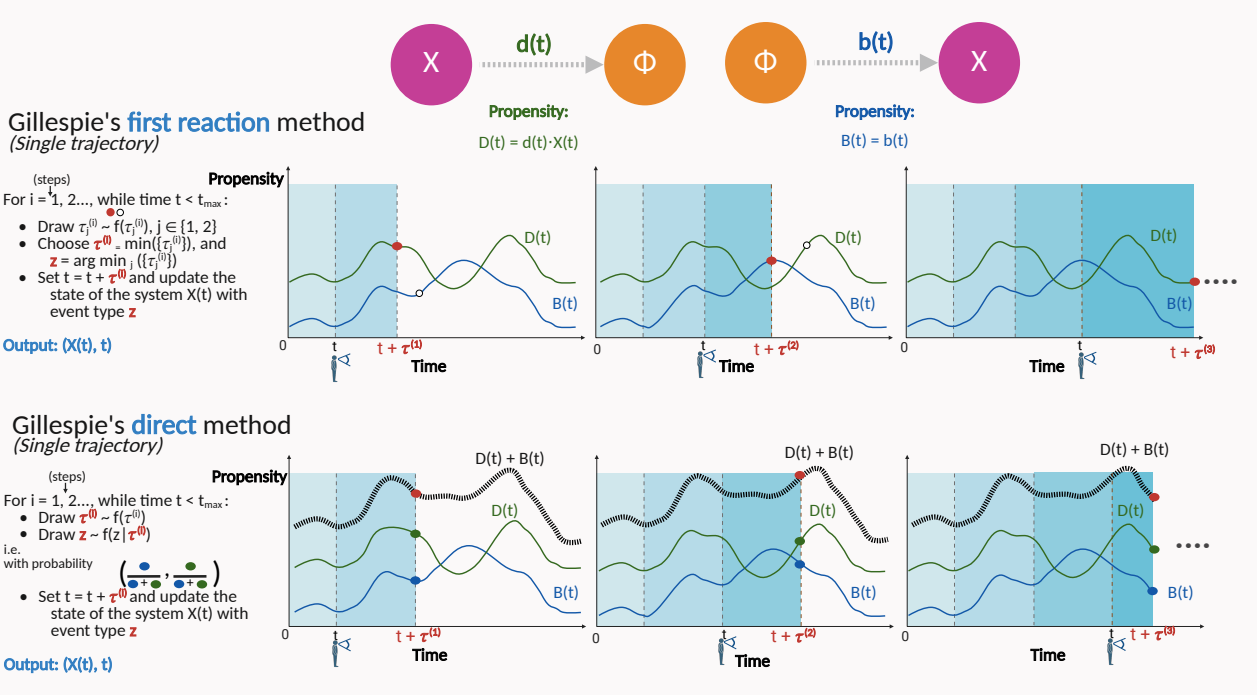


Figure 7.1: **Schematic of our proposed Gillespie simulation algorithms:** The figure visualises a simplified version of our Gillespie direct and first reaction algorithms for time-varying propensities between events (Algorithms (4) and (6) respectively). We focus here on how the algorithms generate trajectories for the simple birth-death system of eq. (7.4.1), where X denotes the chemical species of interest, $X(t)$ is the current state of the system (e.g. number of molecules), $d(t)$ is the death rate at time t, $D(t)$ is the corresponding propensity for death events at time t, $b(t)$ is the birth rate at time t, $B(t)$ is the corresponding propensity for birth events at time t, and Φ is a species no longer of interest. This iterative procedure, for both direct and first reaction algorithms, produces a single trajectory, and works by stepping forward time, in time steps of size τ where the system is changed using the sampled type of event z at each step, from $t = 0$ until a user-defined maximum time or another user-specified condition. Further details on the direct and first reaction methods are included throughout Section 7.3.1 –7.3.2. For implementation of both algorithms via thinning, see Algorithms (3) and (5).

7.3 Gillespie simulation algorithms for time-varying propensities between events

7.3.1 Gillespie's direct method

We consider now where we have propensities that vary over time between events, arising due to external time-varying influences that act on the modelled system between two consecutive event times. For example, our birth or death rates for individual molecules may be subject to time-varying, environmental exposures. Event propensities $a_z(t, X(t))$ are then a function of both time t and the current state of the system $X(t)$. The original Gillespie direct and first reaction methods (Algorithms (1) and (2) respectively) are not directly applicable as the waiting time until the next event or the individual event candidate times are no longer exponentially distributed.

To derive an extension of the direct method (Algorithm (1)), suppose our system is currently in state $X(t)$ at time t and we wish to determine both the time and type of the next reaction event (Figure 7.1). Similar to Section 7.2.1 and Gillespie[51], we first derive $\mathbb{P}_z(t+\tau, t+\tau+dt)$ – the probability that the next reaction will occur in the time interval $(t+\tau, t+\tau+dt)$ and will be a reaction event of type z, where dt denotes an infinitesimally small time step.

Denoting the joint probability distribution of this event as $f(\tau, z)$, we have:

$$f(\tau, z)dt = \mathbb{P}(t' > t+\tau) \cdot \mathbb{P}(\text{type } z \text{ reaction occurs in } (t+\tau, t+\tau+dt)) \tag{7.3.1}$$

$$= \mathbb{P}(t' > t+\tau) \cdot a_z(\tau)dt, \tag{7.3.2}$$

where $\mathbb{P}(t' > t + \tau)$ denotes the survival probability (i.e., that no reaction event occurs in the time interval $[t, t + \tau)$) and $a_z(\tau)$ is the propensity for reaction event of type z during the infinitesimal interval $(t + \tau, t + \tau + dt)$. In our notation, we again drop the dependence of $a_z(\tau)$ on the state of the system $X(t + \tau)$ as $a_z(\tau)$ is fixed with respect to the state of the system during the infinitesimal time interval $(t + \tau, t + \tau + dt)$. This makes clear that the only drivers of time variations in the propensity during this infinitesimal time interval are the external temporal influences on the system, since no events in the system can occur before the first sampled event to alter the propensity through the system's states. For our example, no birth or death events could have altered the number of molecules, and the only time variability in propensities during the time interval dt is from the external time-varying influences acting on the birth and death rates of molecules.

If we have M different types of reaction events, the survival probability that the waiting time until the next event will be greater than τ is:

$$S(\tau) = \mathbb{P}(t' > t + \tau) = \exp\left(-\int_0^\tau \lambda(s)ds\right), \tag{7.3.3}$$

where $\lambda(s) := \sum_{j=1}^{M} a_j(s)$ is the total propensity for waiting time s at time $t + s$, defined as the sum of the propensities for the M different types of events at time $t + s$.

Then, eq. (7.3.1) becomes:

$$f(\tau, z)dt = \exp\left(-\int_0^\tau \sum_{j=1}^{M} a_j(s)ds\right) \cdot a_z(\tau)dt. \tag{7.3.4}$$

Alternatively, we may write this as:

$$f(\tau, z) = f(\tau)f(\tau \mid z) \tag{7.3.5}$$

where the probability density, $f(\tau)$, for the waiting time until the next event τ can be written as:

$$\begin{aligned} f(\tau) &= -S'(\tau) \\ &= -\frac{d}{d\tau}\left[\exp\left(-\int_0^\tau \sum_{j=1}^{M} a_j(s)ds\right)\right] \\ &= \sum_{j=1}^{M} a_j(\tau) \exp\left(-\int_0^\tau \sum_{j=1}^{M} a_j(s)ds\right). \end{aligned} \tag{7.3.6}$$

Then:

$$\begin{aligned} f(z \mid \tau) &= \frac{f(\tau, z)}{f(\tau)} \\ &= \frac{\exp\left(-\int_0^\tau \sum_{j=1}^{M} a_j(s)ds\right) \cdot a_z(\tau)}{\sum_{j=1}^{M} a_j(\tau) \exp\left(-\int_0^\tau \sum_{j=1}^{M} a_j(s)ds\right)} \\ &= \frac{a_z(\tau)}{\sum_{j=1}^{M} a_j(\tau)} \\ &= \frac{a_z(\tau)}{\lambda(\tau)}. \end{aligned} \tag{7.3.7}$$

Importantly, from eq. (7.3.7), conditional on waiting time τ, determining the type of the next event only uses information at reaction time $t + \tau$. There is no dependence on the propensity function before time $t + \tau$.

To derive an algorithm analogous to the original Gillespie direct algorithm[50,51] (Algorithm (1)), we need a routine to sample both i) τ, the waiting time until the next event (eq. (7.3.6)), and ii) z, the type of the next event given waiting time τ (eq. (7.3.7)).

For sampling waiting time τ, we can use inverse transform sampling for only the simplest propensity functions that vary over time. In many cases, the lack of a closed-form inverse CDF F^{-1} means that an alternative approach is required. *Thinning*[626] is one such approach that handles time-varying propensities $a_z(\tau)$.

For sampling waiting time τ, thinning works by the user first selecting an upper bound (or maximum) for the time-varying total propensity $\lambda(\tau)$. This upper bound is an important algorithm design choice as it strongly influences the efficiency of the thinning algorithm by determining its acceptance rate (the proportion of sampled times that are accepted). The selected bound will likely need to change over time to i) remain accurate – it needs to be above possible propensity values, and ii) remain efficient – excessively large bounds could be prohibitively inefficient.

This suggests the following thinning algorithm for sampling the waiting time until the next event τ:

Algorithm 3 Thinning algorithm for Gillespie's direct method

Given the current state of the system $X(t)$ at time t:

1: Compute propensities $a_j(0)\ \forall\ j \in \{1, \dots M\}$, total propensity $\lambda(0) = \sum_{j=1}^{M} a_j(0)$ and set $\tau = 0$
2: Compute upper bound $\lambda_{\max}$ on total propensity
3: Draw waiting time: $\tau_{\text{new}} \sim \text{Exponential}(\lambda_{\max})$
4: Set $\tau = \tau + \tau_{\text{new}}$
5: Compute proposed propensity: $\lambda_{\text{proposed}} = \sum_{j=1}^{M} a_j(\tau)$
6: Compute acceptance probability: $r = \frac{\lambda_{\text{proposed}}}{\lambda_{\max}}$
7: Draw $u \sim \text{Uniform}(0, 1)$
8: **while** $r < u$ **do**
9: Repeat steps 2 – 7
10: **end while**
11: **Output:** τ

For computing the type of next event z, from eq. (7.3.7), we need to sample z with probability $\frac{a_z(\tau)}{\lambda(\tau)}$. Similar to eq (7.2.8) for the fixed event propensity, this can be done as follows:

$$z \sim \text{Categorical}\left(\frac{a_1(\tau)}{\lambda(\tau)}, \frac{a_2(\tau)}{\lambda(\tau)}, \dots \frac{a_M(\tau)}{\lambda(\tau)}\right). \tag{7.3.8}$$

All of this suggests the simulation approach given by Algorithm 4:

Algorithm 4 Gillespie's direct algorithm for time-varying propensities

Initialise time $t = t_0$ and the current state of the system $X(t) = X_0$

1: Draw waiting time $\tau \sim f(\tau)$ using Algorithm (3)
2: Draw event type $z \sim f(z \mid \tau)$ using eq. (7.3.8)
3: Set $t = t + \tau$ and change the state of the system $X(t)$ according to event type z
4: Record $X(t)$. Return to Step 1, or end the simulation
5: **Output:** $(\mathbf{X(t)}, \mathbf{t})$

In each iteration of Algorithm (4), we step time forward by time τ and change the state of the system $X(t)$ using event type z. In our case study, this event is a birth or death of molecule that increases or decreases the population by one, respectively. The trajectory of the system over time is produced by repeating the iterative steps 1–4 until a pre-defined maximum time or pre-specified condition (e.g., population depletion).

7.3.2 Gillespie's first reaction method

We can also develop an algorithm that extends Gillespie's first reaction method (Algorithm (2)). Similar to Section 7.2.2, we again need to first sample candidate waiting times τ_j for each of the M event types. Suppose we are again at time t and $f(\tau_j)$ is the probability density for the waiting time of the next reaction for event type j, defined as:

$$f(\tau_j) = \exp\left(-\int_0^{\tau_j} a_j(s)ds\right) \cdot a_j(\tau_j). \tag{7.3.9}$$

Let $\tilde{f}(\tau, z)dt$ be the probability that the first reaction method produces a reaction of type z in the interval $(t+\tau, t+\tau+dt)$. This probability can be written as:

$$\tilde{f}(\tau, z)dt = \underbrace{\mathbb{P}(\tau < \tau_z < \tau + dt)}_{(a)} \cdot \underbrace{\mathbb{P}(\tau_z > \tau, \forall j \neq z)}_{(b)}, \tag{7.3.10}$$

where, for (b), we have used the assumption that two reactions occurring in $(t+\tau, t+\tau+dt)$ is of order zero.

Considering (a) in eq. (7.3.10), we have:

$$\mathbb{P}(\tau < \tau_z < \tau + dt) = \exp\left(-\int_0^{\tau} a_z(s)ds\right) a_z(\tau)dt, \tag{7.3.11}$$

while (b) in eq. (7.3.10) can be written as:

$$\begin{aligned}\mathbb{P}(\tau_z > \tau, \forall j \neq z) &= \prod_{\substack{j=1 \\ j\neq z}}^{M} \exp\left(-\int_0^{\tau} a_j(s)ds\right) \\ &= \exp\left(-\int_0^{\tau} \sum_{\substack{j=1 \\ j\neq z}}^{M} a_j(s)ds\right),\end{aligned} \tag{7.3.12}$$

where we have used the independence of events at a single iteration of the simulation routine.

We can now rewrite eq. (7.3.10) as:

$$\tilde{f}(\tau, z)dt = a_z(\tau) \exp\left(-\int_0^{\tau} \sum_{j=1}^{M} a_j(s)ds\right)dt, \tag{7.3.13}$$

which coincides with eq. (7.3.4) – the joint probability that the next reaction occurs in $(t+\tau, t+\tau+dt)$ and will be of type z is the same under the direct and first reaction methods.

As part of the first reaction algorithm, we can again use thinning in a similar procedure to Algorithm (3) for sampling individual candidate event times τ_j:

Algorithm 5 Thinning algorithm for Gillespie's first reaction method

Given the current state of the system $X(t)$ at time t:

1: Compute propensities $a_j(0)$ for event type $j \in \{1, \dots M\}$ and set $\tau_j = 0$
2: Compute upper bound $\lambda_{\max}$ for propensity of event type j
3: Draw waiting time: $\tau_{\text{new}} \sim \text{Exponential}(\lambda_{\max})$
4: Set $\tau_j = \tau_j + \tau_{\text{new}}$
5: Compute proposed propensity: $\lambda_{\text{proposed}} = a_j(\tau)$
6: Compute acceptance probability: $r = \frac{\lambda_{\text{proposed}}}{\lambda_{\max}}$
7: Draw $u \sim \text{Uniform}(0, 1)$
8: **while** $r < u$ **do**
    Repeat steps 2 – 7
9: **end while**
10: **Output:** τ_j

All of this suggests the simulation approach given by Algorithm 6:

Algorithm 6 Gillespie's first reaction algorithm for time-varying propensities

Initialise time $t = t_0$ and the current state of the system $X(t) = X_0$

1: Draw each candidate waiting time $\tau_j \sim f(\tau_j) \; \forall \; j \in \{1, \dots M\}$ using Algorithm (5)
2: Choose waiting time $\tau = \min\left(\{\tau_j\}\right)$ and event type $z = \arg\min_j\left(\{\tau_j\}\right)$
3: Set $t = t + \tau$ and change the state of the system $X(t)$ according to event type z
4: Record $X(t)$. Return to Step 1, or end the simulation
5: **Output:** $(\mathbf{X(t)}, \mathbf{t})$

Comparing the direct and first reaction algorithms: i) Both algorithms do not require propensities that are integrable with respect to time, ii) both algorithms could struggle from computational inefficiency from thinning, suggesting the need for analyses of the impacts of adaptive bounds/maxima, iii) the first reaction algorithm may be more computational inefficient as it again requires M different candidate times, all sampled using potentially inefficient thinning, and iv) the first reaction algorithm may be simpler to implement computationally as it does not require computation of $f(z \mid \tau)$.

7.4 Applications of algorithms to chemical reaction systems with time-varying propensities

For our applications, we focus on simulation for systems with time-varying propensities to demonstrate and validate our new simulation methods (Algorithms (4) and (6)).

7.4.1 Problem setup

Following our case study, we will first apply the algorithms to a simple system of two chemical reaction processes, where the $M = 2$ types of event represent molecule production/birth and degradation/death events, respectively:

We set up this problem as:

$$X \xrightarrow{d(t)} \Phi, \quad \Phi \xrightarrow{b(t)} X, \tag{7.4.1}$$

where X represents the chemical species of interest, $d(t)$ is the death rate of an individual molecule at time t, $b(t)$ is the birth rate of an individual molecule at time t, and Φ is a species no longer of interest.

Writing $P_n(t) = \mathbb{P}(X(t) = n)$ to denote the probability distribution for the population in state n at time t, we set up boundary conditions:

$$P_K(0) = 1, \tag{7.4.2}$$

$$P_{n<K}(0) = 0, \tag{7.4.3}$$

which leads to the following system of ordinary differential equations (ODEs):

$$\frac{dP_n}{dt} = (n+1)d(t)P_{n+1} - nd(t)P_n + b(t)P_{n-1} - b(t)P_n \tag{7.4.4}$$

$$\frac{dP_0}{dt} = d(t)P_1 - b(t)P_0, . \tag{7.4.5}$$

where each population state n is governed by its own ODE which is coupled with ODEs for other population states. Unlike the few simple stochastic processes where it is available, the chemical master equation (CME) cannot be solved analytically. We focus on examples where the CME is not available analytically as this is a primary reason for using stochastic simulation algorithms[615,618].

Here, we instead use the set of coupled ODEs (eqs (7.4.4) – (7.4.5)), with n in numerical applications ranging from 1 to a suitably large bound/maximum population value, $n_{\max}$.

7.4.2 Numerical ODE simulation

To empirically verify the Gillespie algorithms derived in Algorithms (4) and (6), we numerically simulate the system described by eqs (7.4.2) – (7.4.5). For a numerical solver of the ODEs, we used the deSolve package in R version 4.4.3[584,627]. We applied the LSODA method (Livermore Solver for ODEs with Automatic method switching)[628,629], an adaptive numerical solver that automatically switches between stiff and non-stiff systems.

We implemented deSolve for the systems (see below for assumptions of $b(t)$ and $d(t)$) from times 0 to 100 at time points spaced at intervals of length 0.5. From these numerical ODE solutions, we calculated the mean and variance of the trajectories at individual time points t as follows:

$$\mathbb{E}[X(t)] = \sum_{n=0}^{n_{\max}} n \cdot P_n(t) \tag{7.4.6}$$

$$\mathbb{V}[X(t)] = \sum_{n=0}^{n_{\max}} n^2 \cdot P_n(t) - \mathbb{E}[X(t)]^2. \tag{7.4.7}$$

Despite the unavailability of a closed-form analytical solution, we can compare the results of our Gillespie algorithms to those produced by the numerical ODE solver and use eqs (7.4.6) – (7.4.7) for calculating the mean trajectory over time and uncertainty intervals around the mean trajectory.

7.4.3 Time-varying birth and death rates and stochastic simulation details

In a first example for the system described by eqs (eqs (7.4.2) – (7.4.4), suppose we have exponentially declining birth and death rates for individual molecules, denoted $b(t)$ and $d(t)$ respectively:

$$b(t) := c_1 \exp(-k_1 t), \tag{7.4.8}$$

$$d(t) := c_2 \exp(-k_2 t), \tag{7.4.9}$$

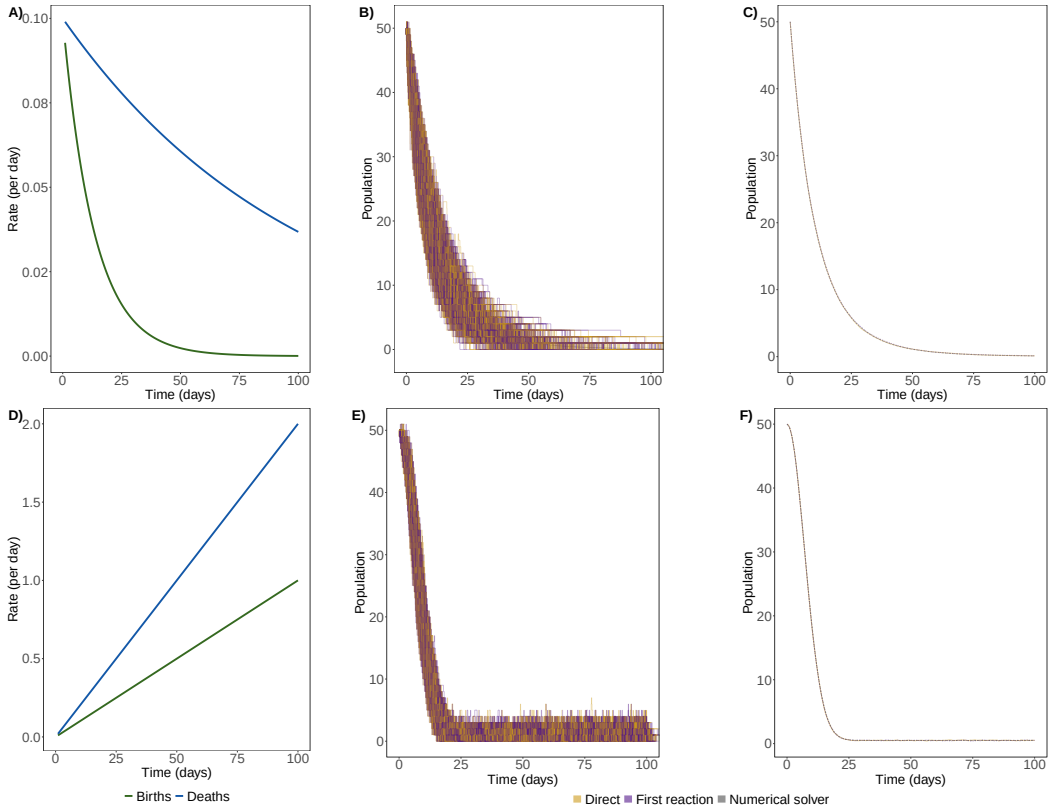


Figure 7.2: **Results for simple chemical reaction systems:** For chemical reaction systems described by eqs (7.4.2) – (7.4.5), we illustrate both the assumed birth and death rates ($b(t)$ and $d(t)$ respectively) for individual molecules and simulation results under assumptions of the exponentially declining (top row) and linearly increasing (bottom row) birth and death rates ($b(t)$ and $d(t)$ respectively). The exponentially declining rates are defined using eqs (7.4.8) – (7.4.9) with $c_1 = c_2 = 0.1$, $k_1 = 0.075$, and $k_2 = 0.01$. The linearly increasing rates are defined using defined using eqs (7.4.12) – (7.4.13) with $l_1 = 0.01$ and $l_2 = 0.02$. **A)** and **D)** capture the corresponding rates at which births and deaths happen over time, under these exponential and linear environments respectively. **B)** and **E)** capture the results of the individual simulations (from Gillespie algorithms and the numerical ODE solver) using step functions for the exponential and linear environments respectively, while **C)** and **F)** capture the interpolated mean values across these individual simulations for both methods. See main text for more details.

where in simulations (Figure 7.2 A), we set $c_1 = c_2 = 0.1$, $k_1 = 0.075$, and $k_2 = 0.01$. These birth and death rates for molecules are combined with the current state of the system $X(t)$ to determine the event propensities for births and deaths as follows:

$$B(t) := b(t), \tag{7.4.10}$$

$$D(t) := d(t)X(t), \tag{7.4.11}$$

In a second example, we consider the same system (eqs (7.4.2) – (7.4.4)), but use linearly increasing birth and death rates, denoted $b(t)$ and $d(t)$ respectively:

$$b(t) = l_1 t, \tag{7.4.12}$$

$$d(t) = l_2 t, \tag{7.4.13}$$

where in simulations, we set $l_1 = 0.01$ and $l_2 = 0.02$.These functions are combined with the current state of the system $X(t)$ to determine the propensities for births and deaths as follows:

$$B(t) := b(t), \tag{7.4.14}$$

$$D(t) := d(t)X(t), . \tag{7.4.15}$$

We assumed a maximum time of maximum time of $t_{\max} = 100$, a starting population of $K = 50$ and a maximum population $n_{\max} = 400$ for eqs (7.4.6) – (7.4.7).

We simulated $\zeta = 2,500$ trajectories for each of these systems using both the direct and first reaction algorithms (Algorithms (4) and (6), respectively).

As reaction events occur at continuous times, to compare trajectories between the algorithms and with the numerical ODE solver, we interpolated trajectories from time 0 to 100 at time points spaced at intervals of length 0.5. To characterise the typical trajectory behaviour of stochastic simulations, for each algorithm, we then calculated the mean and seven quantiles (0.025, 0.1, 0.25, 0.5, 0.75, 0.9, 0.975) of the interpolated trajectories at each time point across the $\zeta = 2,500$ trajectories. We also calculated the standard deviation of these 2,500 interpolated trajectories at each time point to compare trajectory variability for the algorithms with the variability from the numerical ODE solver.

7.4.4 Results of applications

Focusing on the exponentially declining rates (eqs (7.4.8) – (7.4.9), Figure 7.2 A, Figure SI F.1), the simulated trajectories of populations from the first reaction and direct methods coincided with each another and with the numerical ODE solver (Figure 7.2 B). The mean population values for each Gillespie algorithms' trajectories (Figure 7.2 C) at each time point always aligned with one another and with the numerical ODE solver. Similarly, the distribution of population values from the Gillespie algorithms at each time point closely aligned with each other over time, as summarised in Figure SI F.2 for the seven quantiles of the algorithms' trajectories. Finally, at each time point, the variability in the population values of the simulated trajectories (Figure 7.3 A and Figure SI F.1) from each Gillespie algorithm coincided with the variability from the trajectories of the numerical ODE solver.

We repeated this analysis for the linearly increasing rates (eqs (7.4.8) – (7.4.9), Figure 7.2 D). For both Gillespie algorithms, we again found agreement of the mean (Figure 7.2 F) and variability (Figure 7.3 B, Figure SI F.3) of trajectories at each time point with those produced by the numerical ODE solver. The distribution of population trajectories, summarised by quantiles, was again similar at each time point for both stochastic algorithms (Figure SI F.4).

7.5 Individual-based model of dengue transmission

The simulation methods in this manuscript were motivated by our need to simulate an individual-based model of dengue transmission across humans and mosquitoes. In this system, the frequency at which different events occur can i) vary over time due to external influences, ii) vary across individuals (humans and mosquitoes), and iii) depend on the properties (e.g., counts) of other populations. This means that propensities between events can vary over time.

Dengue is an infectious disease caused by the dengue virus that is spread by female *Aedes* mosquitoes. The reproduction and development of the mosquito and virus can depend on temperature, while mosquitoes require sufficient precipitation to lay eggs[80,82,87,88]. These environmental effects, can be modulated by local time- and space-varying human influences (e.g., water storage practices)[45,78,79] and biological influences (e.g., mosquito thermoregulation, genetics, and adaptation[154,155]). To understand the impacts of stochasticity and time-varying effects on the overall dynamics of a dengue outbreak, it is often useful to model infection processes at the level of individual mosquitoes and humans. This can then inform our understanding of epidemics at the population level.

7.5.1 Problem setup

Consistent with the notation of Mills et al.[48], we create an individual-based model where overall populations of mosquitoes and humans have sub-populations of: i) susceptible, exposed and infectious mosquitoes, denoted S^M, W and V respectively,

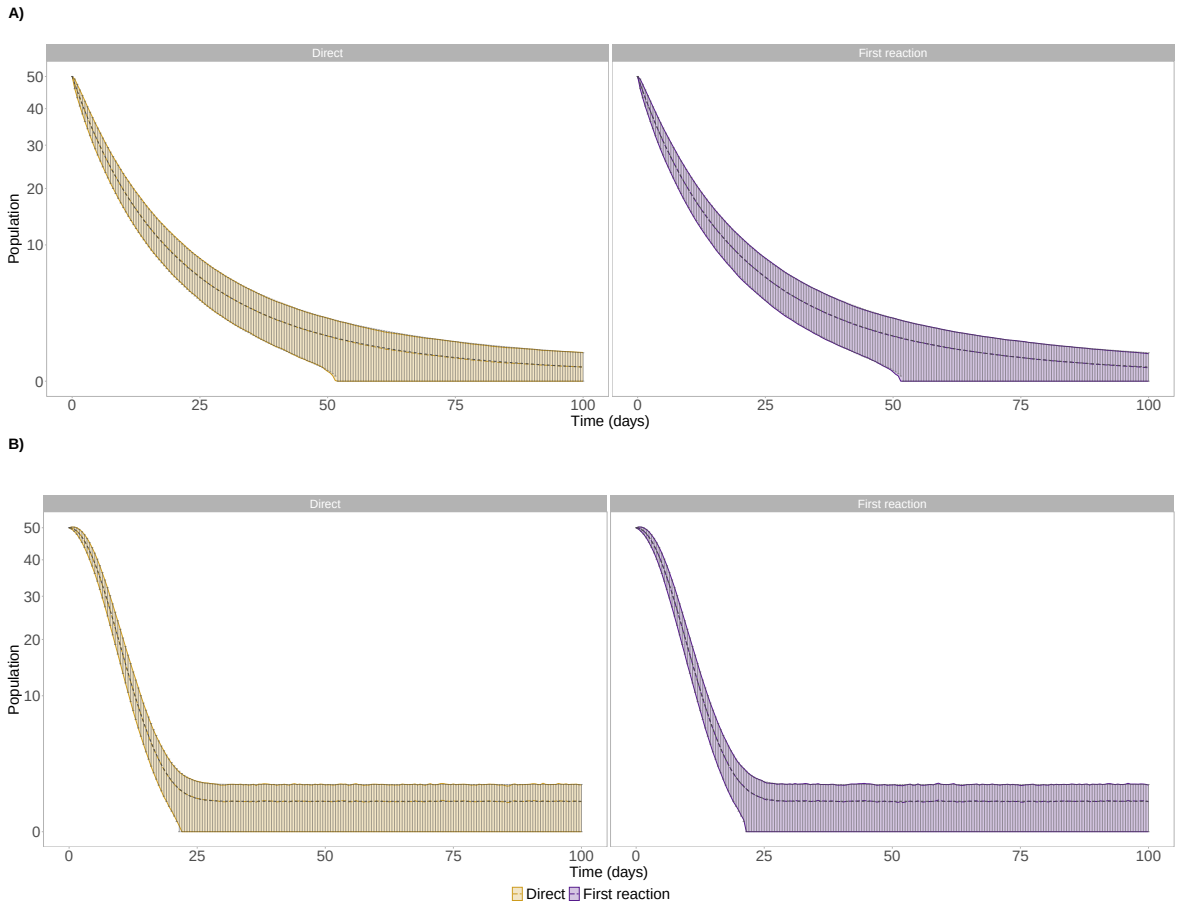


Figure 7.3: **Variability of trajectories for simple chemical reaction systems:** For chemical reaction systems described by eqs (7.4.2) – (7.4.5), we illustrate further simulation results under assumptions of the exponentially declining (top row, **A**)) and linearly increasing (bottom row, **B**)) birth and death rates ($b(t)$ and $d(t)$ respectively). The exponentially declining rates are defined using eqs (7.4.8) – (7.4.9) with $c_1 = c_2 = 0.1$, $k_1 = 0.075$, and $k_2 = 0.01$. The linearly increasing rates are defined using eqs (7.4.12) – (7.4.13) with $l_1 = 0.01$ and $l_2 = 0.02$. The black error bars show uncertainty intervals for the trajectory of the system from the numerical ODE solver, where we used the mean ± one standard deviation. Analogous uncertainty intervals are shown for the interpolated trajectories at each time point from the Gillespie first reaction (purple) and direct (gold) algorithms.

and ii) susceptible, exposed, and infectious humans, denoted S^H, E and I respectively. These sub-populations consist of individuals, each of whom have *ages* (a_j for sub-population j) at each time t. These ages represent the time since the individual entered their current sub-population and do *not* generally represent biological age. Φ denotes all individuals (mosquitoes and humans) no longer relevant to the processes of dengue infection. We use lower case letters to represent individuals in a sub-population. We provide a table of all symbols in Table SI F.1 and assumptions for various parameters in Appendix Section F.2.2.

For simplicity, we do not model many important features of dengue outbreaks including serotype dynamics (e.g., human waning immunity or cross-protection), mosquito population limits (e.g., carrying capacity), or mosquito behaviour and dynamics (e.g., feeding preferences and genetics).

Our individual-based model can be described by the following equations that summarise the competing events:

$$\frac{d}{dt}\left(\Phi \rightarrow S^M\right) = \lambda^M(t) \tag{7.5.1}$$

$$\frac{d}{dt}\left(s^M(t, a_s^M) \rightarrow w\right) = b \cdot \frac{|I(t)|}{N^H} \cdot p_{WI} \tag{7.5.2}$$

$$\frac{d}{dt}\left(w(t, a_w) \rightarrow v\right) = \lambda^W_{\text{Latent}}(a_w) \tag{7.5.3}$$

$$\frac{d}{dt}\left(s^M(t, a_s^M) \rightarrow \Phi\right) = \frac{d}{dt}\left(w(t, a_w) \rightarrow \Phi\right) = \frac{d}{dt}\left(v(t, a_v) \rightarrow \Phi\right) = \mu^M \tag{7.5.4}$$

$$\frac{d}{dt}\left(s^H(t, a_s^H) \rightarrow e\right) = b \cdot \frac{|\ S^H\ |}{N^H}|V(t)| \cdot p_{EV} \tag{7.5.5}$$

$$\frac{d}{dt}\left(e(t, a_e) \rightarrow i\right) = \lambda^E_{\text{Latent}}(a_e) \tag{7.5.6}$$

$$\frac{d}{dt}\left(i(t, a_i) \rightarrow \Phi\right) = \mu^I, \tag{7.5.7}$$

where $\lambda^M(t)$ is the time-varying propensity for susceptible mosquitoes (Figure SI F.5), b is the fixed biting rate of an individual mosquito per day, $|I(t)|$ is the number of infectious humans at time t, N^H is the fixed total number of humans, p_{WI} is the probability of a mosquito becoming infected upon biting an infectious human, $\lambda^W_{\text{Latent}}(a_w)$ is the rate at which an exposed mosquito of age a_w becomes infectious, μ^M is the age- and infection-status-independent mosquito death rate, $|V(t)|$ is the number of infectious mosquitoes at time t, p_{EV} is the probability of a human becoming infected upon being bitten by an infectious mosquito, $\lambda^E_{\text{Latent}}(a_e)$ is the rate at which an exposed human of age a_e becomes infectious, and μ^I is the rate at which an infectious human exits the infectious state. Only the event propensity for susceptible mosquito births is assumed to be time-varying (Figure SI F.5 and Figure 7.4).

We assumed small counts for both human and mosquito sub-populations and that the outbreak started with multiple infected and infectious mosquitoes and one infected and one infectious human. These assumptions form the basis for the following initial populations:

$$X_0(t) = \left(|\ S^M(0)\ |= 25, |\ W(0)\ |= 5, |\ V(0)\ |= 5, |\ S^H(0)\ |= 50, |\ E(0)\ |= 1, |\ I(0)\ |= 1\right), \tag{7.5.8}$$

where we used small population counts as these often create the need for stochastic epidemic models.

If we assume that all rates are independent of age, a *similar* deterministic ODE system for the sub-populations can be written such that:

$$\frac{dS^M}{dt} = \lambda^M(t) - b \cdot \frac{I}{N^H} \cdot S^M \cdot p_{WI} - \mu^M \cdot S^M \tag{7.5.9}$$

$$\frac{dW}{dt} = b \cdot \frac{I}{N^H} \cdot S^M \cdot p_{WI} - \overline{\lambda^W_{\text{Latent}}} \cdot W - \mu^M \cdot W \tag{7.5.10}$$

$$\frac{dV}{dt} = \overline{\lambda^W_{\text{Latent}}} - \mu^M \cdot V \tag{7.5.11}$$

$$\frac{dS^H}{dt} = -b \cdot V \cdot \frac{S^H}{N^H} \cdot p_{EV} \tag{7.5.12}$$

$$\frac{dE}{dt} = b \cdot V \cdot \frac{S^H}{N^H} \cdot p_{EV} - \overline{\lambda^E_{\text{Latent}}} \cdot E \tag{7.5.13}$$

$$\frac{dI}{dt} = \overline{\lambda^E_{\text{Latent}}} \cdot E - \mu^I \cdot I, \tag{7.5.14}$$

$$\tag{7.5.15}$$

where each equation describes rates of changes of the sub-populations and $\overline{\lambda^W_{\text{Latent}}}$ and $\overline{\lambda^E_{\text{Latent}}}$ denote the mean rates (across ages) at which exposed mosquitoes and humans respectively become infectious. We assumed the same initial populations for this system.

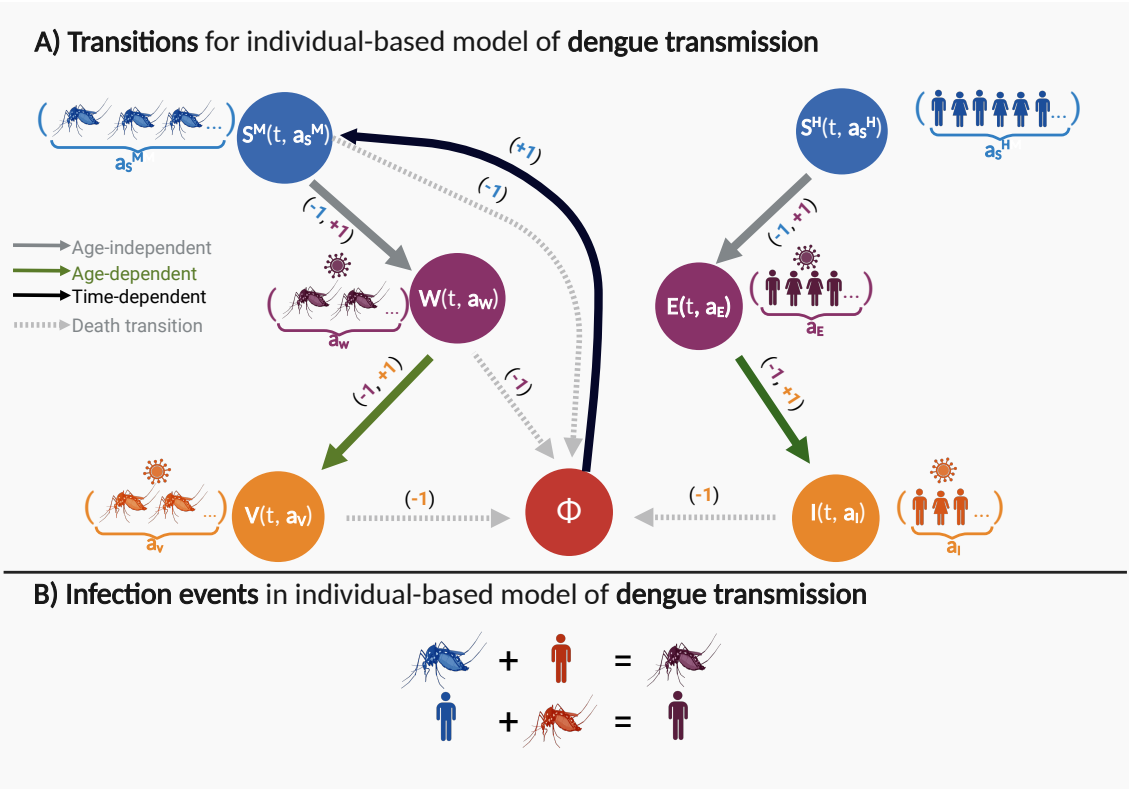


Figure 7.4: **Transitions and infection events for individual-based model: A)** The panel visualises the transitions between sub-populations for the individual-based system of dengue transmission. For a sub-population J, age vector $\mathbf{a_J}$ denotes the ages a_j for individuals in population J at time t, where age denotes the time since the individual entered state j. The evolution of sub-population J is therefore at individual times is described by vector $J(t, \mathbf{a_J})$. The full system is described by eqs (7.5.1) – (7.5.8), while a table of notation is provided in Table SI F.1. **B)** A simplified visual aid for readers from outside infectious disease modelling. To become infected, an individual susceptible mosquito (human) must come into contact with an infectious human (mosquito).

We introduce this similar ODE system as a guide to investigating the behaviour of simulated trajectories for sub-populations eqs (7.5.1) – (7.5.8). We again solve this system of ODEs using deSolve, yet unlike the simple chemical reaction systems of Section 7.4 with coupled ODEs ((7.4.2) – (7.4.5)), the moments (e.g., mean behaviour) of simulated trajectories from a Gillespie algorithm will not necessarily coincide with those from this deterministic ODE system. This deterministic ODE system does *not* capture the probability distribution for each of the sub-population states (see Discussion) which would require a high-dimensional coupling of ODEs across the many states of each sub-population.

7.5.2 Application details and results

In this exploratory example for dengue transmission, we only consider Gillespie's first reaction method (Algorithm (6) to avoid the computation of $f(z \mid \tau)$ for the many types of events (Figure 7.4). This is because the first reaction method, while more computationally intensive, is easier to implement. In our simulations, each individual mosquito and human is subject to one or more type of reaction event process. For example, an exposed mosquito may become infectious or may die. This means that the number of event types varies over time as sub-population counts evolve. We used $\zeta = 100$ trajectories because each trajectory took approximately 1.75 hours to compute.

Using the first reaction method for time-varying propensities, the stochastic mean behaviour of trajectories for the various sub-populations captures an outbreak which peaks fast and declines gradually over a longer time period up to the maximum time of 75 (Figure 7.5). Individual trajectories, while having little meaning on their own, indicate the high

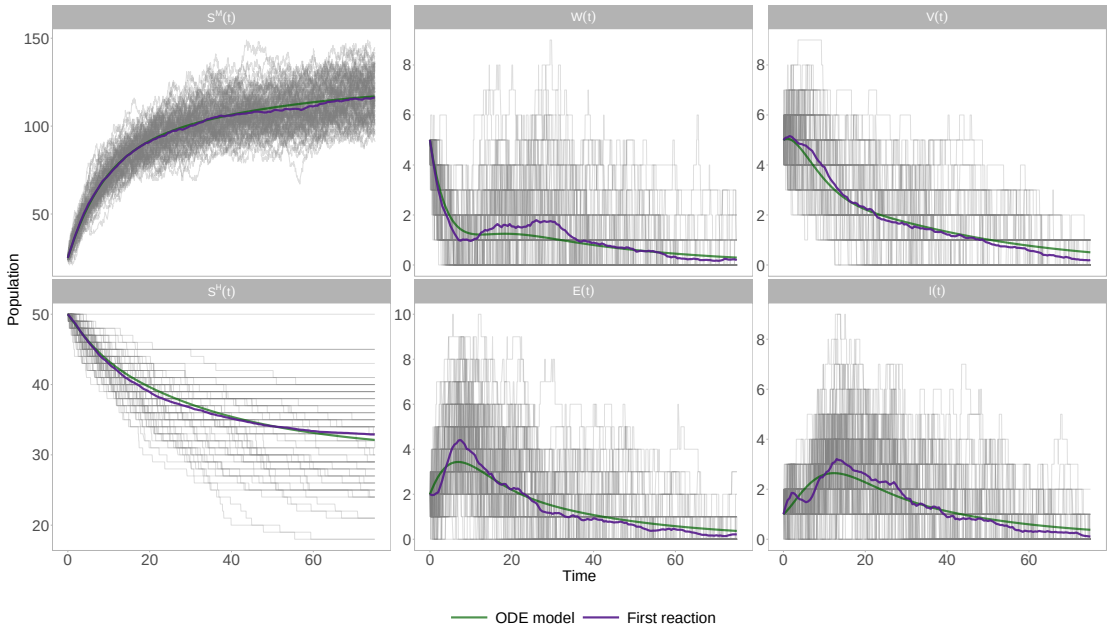


Figure 7.5: **Results for individual-based model of dengue:** Mosquito (top) and human (bottom) sub-populations in $\zeta = 60$ individual trajectories (shaded grey) over time produced by using Gillespie's first reaction method (for time-varying propensities, see Algorithm (6)), alongside the corresponding stochastic mean of sub-population values (purple) at interpolated values from times 0 to 75 days in intervals of length 0.5 days. We include the trajectory from a similar ODE model (green) with age-independent properties (eqs (7.5.9) – (7.5.14)). See Table SI F.1 for a list of all symbols used in this analysis.

degree of variability in the trajectories of individual simulations.

The stochastic mean behaviour of the trajectories from the first reaction method often, but not necessarily, aligned with the trajectory from the ODE model (Figure 7.5). In the absence of simulating from a numerical solver for the probability distributions of each sub-population state, this provides a rough, qualitative safeguard to ensure that our code implementation is not incorrect. Differences over time, especially in the early stages of the outbreak, are also informative since they highlight the possibly high importance of capturing individual-level effects in both human and mosquito populations. These individual-level effects are created by humans and mosquitoes of different ages within each sub-population, and control the rates at which the individual transitions to another sub-population.

7.6 Discussion

In many stochastic systems, from infectious disease transmission to predator-prey interactions, the propensities that control the stochastic reaction rates of the process can vary over time between two events. These time-varying propensities are a function of both the current state of the system $X(t)$ and time t. This situation is caused by time-varying influences that are external to the system studied. External influences can be interacting biological, human, or environmental influences for the availability of prey for predator birth rates or the amount of precipitation for mosquito birth rates. The widely used Gillespie direct and first reaction algorithms[50,51] are not directly applicable to these systems with time-varying propensities, as these algorithms assume fixed event propensities between events. Although there are alternative exact (e.g., a modified next reaction method[624]) and approximate (e.g., Gillespie algorithm with adaptive thinning and the CLT[625]) algorithms, it was less clear how best to incorporate time-varying propensities directly into the original Gillespie algorithms. This is important since next reaction methods can be difficult to code, while approximate methods can induce approximation error and numerical instability.

We introduced exact simulation algorithms that build on Gillespie's direct and first reaction algorithms[50,51] to incorporate the time-varying event propensities between events. In simple chemical reaction systems, we verified the accuracy of these methods by comparing the distribution of trajectories produced by our simulation algorithms with numerical solutions for analogous systems of coupled ODEs.

Our algorithms (Algorithms (4) and (6)) are flexible and may be useful for simulating models with time-varying propensities in many systems, including movement ecology, financial markets, or chemical decay. The algorithms were motivated by our need to simulate a complex individual-based model of dengue transmission in human and mosquito populations. Our dengue simulations are illustrative. They are not calibrated against observational data such as mosquito catch and human case data, nor do they include carrying capacities of mosquito populations. The many assumptions should be tested in sensitivity analyses. The comparison to a similar ODE system without age dependence is also simple and remains a qualitative guide to inspect whether the stochastic algorithms produce plausible trajectories. In future work, it would be instructive to simulate from the high-dimensional system of coupled ODEs across the various human and mosquito sub-populations, similar to those used for the simple chemical reaction system (eqs (7.4.2) – (7.4.5)).

However, developing and simulating these individual-based stochastic systems is a stepping stone towards more realistic models for transmission of dengue and other mosquito-borne pathogens. There are numerous environmental sensitivities across the mosquito and virus (e.g., temperature, precipitation, and humidity)[80,84,526]. It is also less well understood how many laboratory-derived results (e.g., temperature-dependent mosquito and viral traits) extrapolate to the field where environmental influences interact with one another and with human and biological influences. Having an individual-based simulator against which we can calibrate a population-level model (e.g.,[48]) is one way to ensure that inference procedures are robust and working properly. Our example is specific, yet is reflective of a more general procedure where individual-based simulations can i) capture fine-grained details (e.g., individual variability and stochastic effects) of a stochastic, non-linear, and complex system, and ii) be paired with population-level models to test the rigour of inference methods for the parameters of the corresponding population-level model.

Our Gillespie algorithms (Algorithms (4) and (6)) adopt a different approach to the modified next reaction method, which uses a propensity queue and Poisson process (see Introduction and[624]). We focus only on extending the traditional Gillespie direct and first reaction methods. We suspect that the modified next reaction method will be more suitable for efficient and exact simulation of large and/or complex systems with potentially many event types. On the other hand, this modified next reaction algorithm is limited by known difficulties in computational implementation and the assumption of an integrable propensity function[615,618]. There are also trade-offs between our algorithms being exact simulation algorithms, instead of approximate methods. Computational inefficiency may mean that other approximate simulation methods (e.g.,[625]) may also be more efficient in high-dimensional systems. These high-dimensional systems include individual-based models of virus transmission across many hosts and vectors, chemical reaction networks with low numbers of molecules, or high-frequency financial markets. However, in such high-dimensional systems, stochastic models may not even be as important as individual-level effects begin to cancel each other out. Due to the same computational inefficiency that Gillespie methods have always faced[615,618], our methods are likely i) most suitable for stochastic simulation of small systems and ii) a theoretical basis for future more efficient algorithms (similar to the classical Gillespie algorithms as the basis for efficient elaborations). These issues are made worse for our algorithms by the need to perform thinning to sample from time-varying reaction rates.

In future works, our simple algorithms could be adapted using efficiency improvements already used when propensities are fixed between events, including τ-leaping methods[621,622,623]. As an alternative scheme, we suspect that a piecewise-constant algorithm could be both easily to implement and computationally useful. Located somewhere between Gillespie's classical algorithms (Algorithms (1) and (2)) and our extended algorithms (Algorithms (4) and (6)), we could adapt a fixed-propensity Gillespie algorithm to only update the external, time-varying component of the event propensities at the

start of each time interval. This proposed approximate algorithm is more efficient than our exact algorithms (Algorithms (4) and (6)) and its accuracy will be controlled by the length of the time interval between events. For short time intervals that often occur in large systems, this approximate method will likely be highly accurate, yet may struggle for longer time intervals and small systems (where our exact algorithms perform best).

Other ways to improve the efficiency of our exact or approximate methods may include vectorisation in code implementations, deep learning approaches (e.g., graph neural network representations with sparsity or emulators such as Gaussian Processes), or other adaptations of optimisation techniques from stochastic process literature (e.g., modifying transition rates[630]).

In conclusion, our methods establish a foundation for the exact simulation of small stochastic reaction systems with external variability, with applications spanning multiple branches of science.

Data Availability Statement

All code and data used in our analysis are available at `https://github.com/cathalmills/gillespie_time_varying/`.

Chapter 8

From metric to action: An evaluation framework to translate infectious disease forecasts into policy decisions

Cathal Mills, Nicholas J. Irons, Joseph L.-H. Tsui, Luiz Max Carvalho, Claudia T. Codeco, Adam J. Kucharski, Oliver Ratmann, Ben Lambert, Christl A. Donnelly, and Moritz U. G. Kraemer. From metric to action: An evaluation framework to translate infectious disease forecasts into policy decisions. June 2025

This chapter presents a manuscript to measure and improve the actionability of infectious disease forecasts[52]. Our new evaluation framework shifts the perspective of forecast evaluation from the forecaster to the local decision-maker. The manuscript will be submitted to *The International Journal of Forecasting*.

Abstract

Decisions during an infectious disease outbreak are often made under considerable time and societal pressure and in the face of great uncertainty. These difficult and consequential decisions can ideally be supported by a probabilistic forecast – the prediction of the future value of an epidemiological quantity, together with uncertainty. Currently, forecast performance metrics generally focus on the forecaster's perspective; there is a lack of principled protocols for forecast evaluators to measure a forecast's "value" – its ability to provide actionable insights for decision-makers. We develop an evaluation framework for improved decision-making, focusing on three aspects: i) translating popular forecast evaluation metrics into interpretable, public-health-relevant quantities across space and time; ii) converting statistical predictions and evaluations into actionable outputs; iii) linking predictability of an epidemic to the value of forecasts for decision-making. By melding concepts from weather forecasting, information theory, and decision theory, the framework provides tools that allow forecasters and forecast evaluators to focus on the decision-maker, considering their priorities and risk preferences in different settings. Throughout, we highlight how the limits of epidemic predictability impact the value of a forecast, particularly vis-à-vis the decision-maker's dilemma: to act or not. We illustrate the framework with an application to forecasts of weekly incident COVID-19 cases. Our evaluation framework connects statistical concepts with decision-theoretic principles to promote more actionable and reliable infectious disease forecasts that enhance public health responses.

- Conceptualisation: **C.M**; M.U.G.K

- Methodology: **C.M**; M.U.G.K; N.I
- Data curation and visualisation: **C.M**
- Writing original draft: **C.M**; N.I
- Writing final draft: **C.M**; M.U.G.K; N.I; B.L; O.R; A.J.K; C.A.D; L.M.C;J.L.H.T; C.C; S.S

8.1 Introduction

> "Forecasts possess no intrinsic value. They acquire value through their ability to influence decisions made by users of the forecasts [161]."
>
> *Allan H. Murphy*

Infectious disease forecasting involves predicting future values of epidemic indicators over short- and medium-term horizons. These indicators may be disease-induced cases, hospitalisations, healthcare capacity, or deaths, and are distinguished from *scenario projections* by being unconditional predictions about the future. Subseasonal or short-term epidemic forecasts can be used to inform outbreak responses, including when and where to deploy resources, implement interventions, and launch public information campaigns. Seasonal (i.e., medium-term) forecasts can be used to quantitatively inform public health planning, as predicting the season's peak (timing and value) and cumulative cases/hospitalisations/deaths can provide timely information about future resource requirements. Depending on the targeted epidemic indicator and actions of decision-makers, the predicted future can be i) modifiable (e.g., COVID-19 cases following vaccination strategies with direct impacts), or ii) largely invariant to interventions (e.g., seasonal dengue cases following expansion of healthcare capacity). A unifying goal in all forecasting tasks is to support the forecast *user* (i.e., a decision-maker) by providing the scientific understanding and uncertainty necessary for making informed decisions.

Although there were pre-COVID-19 pandemic guidelines [468,564] and forecast competitions for several infectious diseases (e.g., dengue, Ebola, and seasonal influenza [169,171,631]), there has been rapid progress in infectious disease forecasting since the COVID-19 pandemic. Many of the key advances (e.g., the COVID-19 Forecast Hub in the United States [170,173,433]) were in the evaluation of model performance and the construction of ensembles for improved forecast performance. At the heart of such breakthroughs has often been the use of scoring rules (which measure the whole forecast distribution) and more detailed forecast performance evaluations. Such methods are the gold standard for evaluating the reliability of infectious disease forecasts and nowcasts (e.g., as measured by the calibration of their uncertainty intervals) [167,213,366,433,469,473,572,632,633]. Separately, statistical decision theory has been proposed for more principled epidemic decision-making processes [105,160].

Many advances in infectious disease forecast evaluation have focused on forecast *quality* (alignment of probabilistic forecasts with observations), yet fewer examples link infectious disease forecasts to *value* for decision-makers (i.e., benefit to the forecast user to make decisions). In weather forecasting [634] and in medical decision-making [635], there are clearer guidelines for linking model outputs to the value for the decision-maker. Meanwhile, it remains unclear how to i) systematically evaluate infectious disease forecasts for individual decision-makers under considerable uncertainty and time pressure, and ii) convert the statistical evaluation metrics to timely actions. Although there are differences with infectious disease forecasting (e.g., the better understanding of the physics of weather or the inability to influence dynamics) [288], in weather forecasting, evaluation procedures measure how well a model's forecast informs the user's ability to take action for extreme weather events (e.g., weather warnings for impact mitigation by emergency managers or for travel avoidance by the public) [634,636].

Relevant questions within the context of infectious disease forecasting remain understudied. Is a model "good" at producing forecasts that inform important epidemic decisions? Can a decision-maker take appropriate action from statistical metrics and the outputs of a set of forecast models? How do forecast model rankings and decision-makers' decisions vary with different risk preferences? Similarly, to ensure usefulness to decision-makers and safeguard against the erosion of trust, reported forecasting model outputs must be reliable. However, there are no procedures to determine *a priori* which future forecast model outputs are likely to be reliable under different epidemic regimes. Is it possible to anticipate the future reliability of model forecasts to determine which models to use, additional data to collect, or outputs to report?

We address each of these questions by building on recent advances in infectious disease settings and other branches of science. We develop a framework for infectious disease forecasters and forecast evaluators to i) interpret the contributors to the forecast performance of models (Section 8.2), ii) quantify the forecast value for specific epidemic events and decision-makers (Section 8.3), and iii) address the effects of uncertainty and predictability on the forecast performance and value of a model (Sections 8.4 – 8.5). The framework simultaneously emphasises statistical rigour and public health utility throughout, while explicitly recognising practical time pressures and uncertainties of epidemic decision-making.

8.2 Recasting scoring metrics for interpretable epidemic forecasting across space and time

The targeted quantity, model, and evaluation procedure for an infectious disease forecast should depend on the decision-maker and their policy questions and risk preferences. Before the decision-maker can decide whether to take action, both the forecaster and forecast evaluator must understand the overall quality and value of the model's forecasts and their determinants. We will explain ways to determine whether models have stable performance across different statistical properties, decision-makers, and policy questions, and how performance differs across space and time. Rather than interpreting popular scoring metrics as the end goal of evaluation, we focus on the implications and implicit assumptions of metrics for infectious disease forecasting and then discuss how we can extend the application of these concepts and integrate them into an evaluation framework (Figure 8.1 and Section 8.6).

8.2.1 Scoring functions and rules

Probabilistic forecasts of epidemic quantities can be comprehensively evaluated by proper scoring rules. These are loss functions that measure reliability by how well the model captures uncertainty (using the whole forecast distribution) and quantify distributional accuracy beyond the alignment of pointwise estimates with observations. A scoring rule is a specific type of scoring function (see Table 8.1) with lower values implying better performance. A scoring rule is (strictly) proper if the score is minimised, in expectation, by (only) the unknown, true probability distribution of the target quantity. Examples include the continuous rank probability score (CRPS)[470] and the Brier score (BS)[637]. For predictive cumulative distribution F and continuous observation y (e.g., incident hospitalisations), the CRPS is defined as:

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - 1(x \geq y))^2 dx. \tag{8.2.1}$$

Probabilistic forecasts are often represented using quantiles of the predictive distribution F[167,170,172,434], which can be scored on test data using the Weighted Interval Score (WIS)[58], a proper scoring rule that approximates the CRPS. Relative scores (e.g., relative WIS, rWIS[172,638]), while no longer proper[639], are often used to compare forecast performance against a reference model. WIS and CRPS are our framework's proper scoring rules for measuring overall predictive performance (Figure 8.2 B) of a model for forecasting the continuous value of a target epidemiological quantity y.

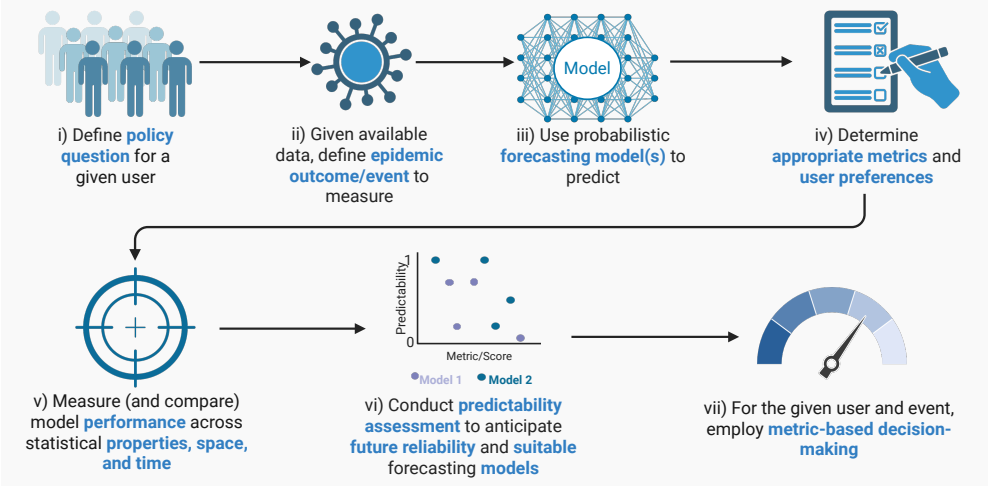


Figure 8.1: **Evaluation framework and workflow for forecasters and forecast evaluators:** A proposed workflow for our evaluation framework which enables forecasters and evaluators to justify why a decision-maker would use the forecasts of a specific model (or set of models) to inform a public health policy decision. The workflow starts with the forecaster or evaluator focusing on a specific policy question co-defined with a given decision-maker. The workflow is to be applied iteratively at each time point (e.g., for multiple events and metrics), and the metric-based optimal model(s) for each decision will be based on one or more chosen metrics. For determining policy questions and an appropriate metric for a given user and policy question, see Figure 8.2 and Table 2. We provide details of metric-based and predictability assessments throughout Sections 8.2–8.5. See Section 8.6 for applications of the framework to COVID-19 case forecast evaluation.

The BS is an *event-specific* (Figure 8.2 B) classification measure that assumes a uniform distribution of decision-maker risk preferences[637]. BS is defined for a binary outcome $y \in \{0, 1\}$ and probabilistic forecast $p \in [0, 1]$ as:

$$\mathrm{BS}(p, y) = (p - y)^2 \tag{8.2.2}$$

where $y = 1$ indicates an event occurring (e.g., incident cases above a threshold θ). The CRPS is the BS integrated over all possible thresholds (θ) for the binary event.

Pinball loss (or quantile score) assesses forecasts for a fixed, *user-specific* (Figure 8.2 B) quantile level τ:

$$\mathrm{PL}_\tau(x, y) = (1\{y \le x\} - \tau)(x - y) = \begin{cases} (1 - \tau)(x - y), & y \le x, \\ \tau(y - x), & y \ge x, \end{cases} \tag{8.2.3}$$

where x is the $\tau-$quantile of a predictive distribution F. Pinball loss is a consistent scoring function (see Table 8.1), as opposed to a scoring rule, since it assigns a score to a given quantile (τ) and not the entire forecast distribution. The CRPS is the pinball loss averaged over all quantile levels.

For any scoring function or rule, the mean score across N test set observations is often used as a summary score to rank the out-of-sample abilities of different forecasting models:

$$\overline{S} = \frac{1}{N} \sum_{i=1}^{N} S_i \tag{8.2.4}$$

where S_i is the score (using, e.g., CRPS) obtained for observation i, calculated using one of eqs (8.2.1)–(8.2.3).

Table 8.1: **Key concepts for our framework of probabilistic infectious disease forecasting:** We define important concepts discussed throughout the text and used in our framework for forecasting infectious diseases. Given the wealth of metrics and tools above, the choice of metric/tool may not be immediately obvious, and may be both user- and event-specific, which motivates our workflow (Figure 8.1 A) starting with the well-defined policy question(s) for a user/decision-maker and corresponding epidemic outcome to measure. See Table SI G.1 for definitions of additional concepts.

Concept	Meaning
Brier Score (BS)	A strictly proper scoring rule defined as the mean squared error for a probabilistic forecast of a binary outcome[637].
Classifiers	A probabilistic classifier produces a probability forecast for a binary outcome.
Consistent	A scoring function is consistent for a given functional $T(P)$ (e.g., quantile level) if the score is minimised when the forecast aligns with $T(P)$ consistent.
Continuous Ranked Probability Score (CRPS)	A proper scoring rule that is strict when the first moment of the forecast distribution is finite, and generalises mean absolute error for probabilistic forecasts of continuous outcomes[470].
Cost-loss (C/L) ratio	In this decision model for binary outcome events, C/L ratio α is the cost (C) of preparation for an adverse event, divided by the avoidable loss (L) if the event occurs without any preparation[161,640]. The optimal decision is to act/prepare if the forecasted probability exceeds this *decision threshold* C/L. Knowing plausible values of the ratio for a given binary event is sufficient for studying forecast model results, and thus for metric-based decision-making.
Discrimination (DSC)	A non-negative component of the decomposition for the mean value of a score, and measures the ability of a model's forecasts to distinguish events versus non-events[641]. Higher values indicate superior discrimination ability.
Event threshold	A cut-off value θ that is used for defining whether a binary outcome occurs or not.
Forecast/predictive distribution	A probabilistic forecast's distribution for the future value of a given quantity, often represented by a parametric distribution, samples from the distribution, and/or quantiles of the distribution.
Forecast horizon	The lead time at which forecasts are made.
Forecast quality	Measure of the alignment of a model's forecasted values with observed values.
Forecast value	Measure of the ability of a model's forecasted values to inform decision-making.
Horizon	The lead time at which a forecast for a future value is produced.
Miscalibration (MSC)	A non-negative component of the decomposition for the mean value of a score, and measures the calibration of a model's forecasts[641]. Lower values indicate superior calibration.
Murphy curve/diagram	For a given model, a curve of mean elementary scores (eq. (8.2.6)) across different C/L ratios (α) or event thresholds (θ). Lower curves indicate better performance, and a Murphy diagram compares Murphy curves for different models.
Pinball loss	Measure of forecast performance for a specific user-defined probability/quantile level. Pinball loss is also known as Quantile Score (QS) or the asymmetric piecewise linear scoring function.
Permutation entropy (PE)	Measure of predictability which works by studying temporal dependencies in both linear and nonlinear time series. PE is a model-free measure that uses Shannon's Entropy to study the frequency of distinct patterns in time series data.
Predictability	The inherent randomness of a system which impacts the extent to which future values can be reliably forecasted.
Proper	A scoring rule is proper if the expected score is minimised by using the true probability distribution.
Scoring function	A function that assigns a score to a given functional (e.g., quantile level). This is not necessarily proper, and can asymmetrically penalise probabilistic forecasts (e.g., pinball loss). Lower values for a scoring function imply better predictive performance.
Scoring rule	A scoring rule (e.g., CRPS) is a specific type of scoring function (see above) that assigns a score to a probabilistic forecast, and lower values for a scoring rule imply better predictive performance.

8.2.2 Stratifications and interpretable scoring metrics

Alongside difficulties in interpretability and actionability for decision-makers, which we discuss below, summary scores can obscure i) variations in forecast performance under different epidemic conditions (e.g., temporal[633], feature[170], or spatial[167] dependencies) and ii) variations in the different statistical components of forecast performance (e.g., calibration versus discrimination abilities). Epidemics unfold across space and time, thereby exhibiting distinct spatial and temporal signatures. In our framework, as in previous studies[46,162,167,170,172,592], all evaluation metrics should be applied to all available test data, alongside stratifications – trailing windows/subsets of data, different geographies, and spatial scales (see applications in Section 8.6). This will help i) identify the time- and space-varying differences in model performance (e.g., in the presence of data distribution shifts), and ii) ensure that model development, model comparison, and decision-making procedures reflect the relevant local context: the decision-maker(s), stage of epidemic, and region(s) in question.

Identifying a "best" forecasting model will always be a subjective decision, determined by the policy question (Figure 8.2 A) of a local decision-maker and available data, which then guides the choice of evaluation metric and model (Figure 8.1). Even when appropriate metrics have been chosen (see Table 8.2) and applied carefully to past forecasts across time and space, it is instructive that forecasters and evaluators uncover the determinants of a model's statistical strengths and weaknesses. This is important for model improvements, enhanced trust, and tackling uncertain future conditions. The statistical properties are the drivers of forecast reliability for public health policy. For example, a naïve, constant model for

short-term forecasts could be perfectly calibrated, yet have little discrimination ability for major surges in cases. Another model could produce precise forecasts, yet be over-confident and have poor calibration.

While rarely used in infectious disease forecasting, scoring function decompositions can shed light on the components of forecast reliability[641,642,643,644,645,646,647]. There are various ways to decompose $\overline{S}$, and we use CORP (Consistent, Optimally binned, Reproducible, and Pool-adjacent-violators)-algorithm-based decompositions[641,647,648]. These methods use nonparametric isotonic regression to decompose $\overline{S}$ for any consistent scoring function (eq. (8.2.4) for e.g., CRPS) into three interpretable components:

$$\overline{S} = \text{MCB} - \text{DSC} + \text{UNC}, \tag{8.2.5}$$

where non-negative quantities MCB, DSC, and UNC denote miscalibration, discrimination, and uncertainty. Lower values of MCB indicate superior calibration (i.e., forecasted probabilities align with observed frequencies), while higher values of DSC indicate superior discrimination ability (mean score of the model relative to a recalibrated but constant forecast). Calibration is a measure of forecast reliability, while discrimination ability is a measure of how well a model can distinguish different epidemic events ahead of time. The uncertainty component measures the irreducible variability of the data.

The CORP-based decomposition (see[641,647,648] for further details) relies on the availability of sufficient test data to perform a forecast recalibration. This decomposition may have limited use in early outbreak stages and in short windows of epidemic data (when tracking temporal variations in forecast performance) due to the scarcity of *relevant* data. Future applications of our framework could consider using other existing decomposition methods (e.g.,[649] or[645]) or developing new decomposition techniques (e.g., using conformal inference[425,426,650,651] in the presence of data distribution shifts).

The CORP decomposition for a scoring function estimates the drivers of statistical performance (measured by the score) for a forecasting model, which are also public-health-relevant properties. Strong discrimination ability is crucial for detecting surges and declines in epidemic curves, while probabilistic forecasts must be well-calibrated to maintain reliable estimates of the uncertain future in epidemics.

In our framework, for scoring functions, we visualise CORP-based discrimination abilities alongside calibration properties (Figure 8.2 B), where DSC (y-axis) is plotted against MCB (x-axis), with points indicating models and lines indicating level curves of constant forecast performance (as measured by the scoring function). We encourage also plotting DSC and MCB for operational baseline/reference models (e.g., a seasonal average model for an endemic disease). Diagnosing any discrimination or calibration issues across space and time, or stratified by features, can provide insights for forecasters about where additional data or bias correction may be required and for evaluators about ensemble model weights and likely suitable models for different policy questions.

8.2.3 Elementary scores and utility functions

Despite the central role of cost-benefit analysis in public health planning, it is not often recognised in infectious disease forecasting that all commonly used summary scores are intricately linked to cost-loss (C/L) ratios. These scores are the output of *weighting* (Figure 8.2 B) an *elementary score* by a utility function for the user's preferences (e.g., risk aversion via their C/L ratio) for different event thresholds (e.g., epidemic events)[652,653].

The error scoring matrix is defined as:

	Event occurs	**Event does not occur**
Action taken	C	C
Action not taken	L	0

C will often represent economic costs for a decision-maker taking a specific preventive action (e.g., increasing intensive care unit (ICU) capacity), and the preventable loss L will be the *additional* burden and public health impact (e.g.,

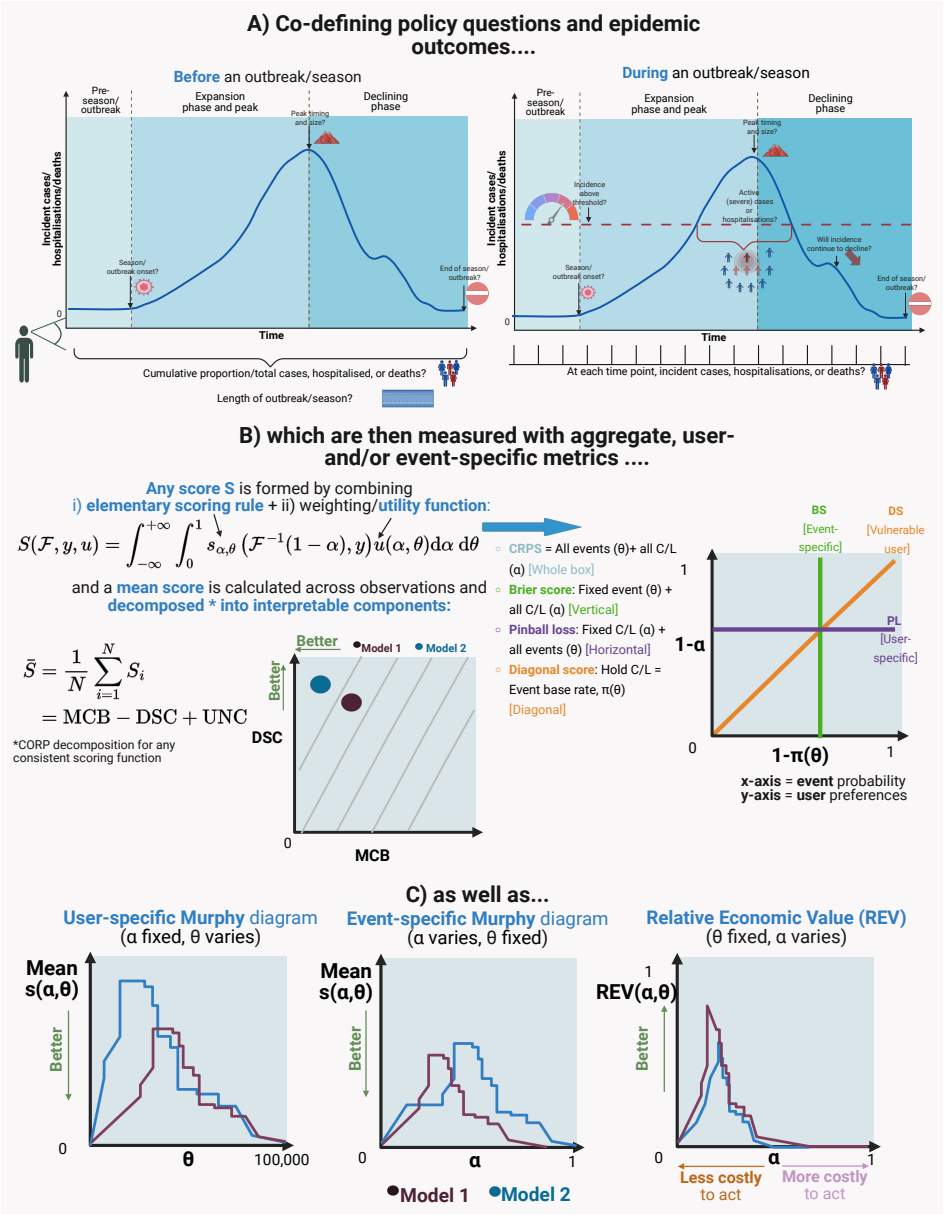


Figure 8.2: **Policy questions and metrics for different users and events: A)** Illustration of generic epidemic curves to illustrate policy questions, either before or during an outbreak/season. All forecast evaluation in our framework starts from these defined policy questions. **B)** Popular summary scores (e.g., CRPS) are formed by combining an elementary score and utility function, and each score makes implicit assumptions about i) the epidemic event which is captured by event threshold θ, and ii) the decision-maker whose preferences for the event are expressed by the C/L ratio α. These scores can be further decomposed into the statistical drivers of forecast performance for enhanced interpretation (see main text). **C)** User- and event-specific evaluation tools – Murphy diagrams and Relative Economic Value curves – where lower and higher curves respectively for a model indicate better forecast performance.

overwhelmed hospital losses, or the more costly loss of lives) if the pre-specified event (e.g., severe outbreak) occurs without taking action. We can also view C as the economic *harm* of action and L as the maximum *benefit* of taking action.

The *elementary score*[652] is defined for an event threshold $\theta \in \mathbb{R}$ (binary event occurs if $y > \theta$) and C/L ratio α (the ratio of the cost of taking action versus the preventable loss if no action taken):

$$s_{\alpha,\theta}(x, y) = \begin{cases} 1-\alpha, & \text{if} \quad x \leq \theta < y \\ \alpha, & \text{if} \quad x > \theta \geq y \\ 0, & \text{otherwise.} \end{cases} \tag{8.2.6}$$

where $x \in \mathbb{R}$ is the $(1-\alpha)-$quantile forecast, $x := F^{-1}(1-\alpha)$.

Under the C/L decision model, $\alpha = C/L$ is the optimal probability threshold above which the decision-maker takes action. Small C/L values for an individual decision-maker indicate inexpensive risk mitigation and large potential losses for the event, and the optimal action for this risk-intolerant decision-maker will be to act even when the forecasted event probability is low.

A C/L ratio can have very different interpretations, depending on how costs and losses are defined (see Section 8.2.4). In general, if $\alpha < 0.5$, the economic costs incurred by an underprediction and lack of action (e.g., too few ICU beds) may be low, yet loss of life could be high as this loss is valued as much more costly than the cost of an ICU bed. Eq. (8.2.6) would penalises this underprediction with a higher elementary score. If $\alpha > 0.5$, the economic costs incurred by an overprediction and prompted action (e.g., too many ICU beds) may be high, relative to the defined losses incurred.

A *summary score* is generally defined by weighting the elementary score (eq.(8.2.6)) with a *utility* function $u(\alpha, \theta)$[652,653].

$$S(F, y, u) = \int_{-\infty}^{+\infty} \int_{0}^{1} s_{\alpha,\theta}(x, y) u(\alpha, \theta) \mathrm{d}\alpha \, \mathrm{d}\theta, \tag{8.2.7}$$

where $x \in \mathbb{R}$ is the forecast at the $(1-\alpha)-$quantile level $x := F^{-1}(1-\alpha)$, $u(\alpha, \theta)$ determines how the forecast user weights different combinations of C/L ratios (i.e., probability decision thresholds, α) and epidemic event thresholds (θ).

Eq. (8.2.7) establishes a key result for both forecasters and evaluators: any summary score must make an assumption about the decision-maker's utility function $u(\alpha, \theta)$. These assumed utility functions each differentially weight, simultaneously, the decision-maker's risk aversion α and epidemic event severity θ. For example, the CRPS (eq. (8.2.1)) is defined by uniformly weighting all probability decision thresholds α and considering all epidemic event thresholds θ. This evaluation stretches across the whole probability space of decision-makers and events (Figure 8.2 B), and CRPS produces the simplest single measure of forecast value for the entire uniform family of decision-maker risk preferences for all epidemic events[654]. While the CRPS provides this rigorous assessment across decision-makers and epidemic events, such implicit assumptions may be less suitable for specific decision-makers and in a fast-evolving epidemic.

To illustrate Eq. (8.2.7), suppose an individual is considering a probabilistic weather forecast (the model) of daily rainfall exceeding a defined level (the event threshold, θ). Their decision problem is whether to bring an umbrella (the action) based on combining the forecast with their risk preferences (α) about the rainfall event threshold (θ). The value of a forecast will clearly depend on the model *and* how the individual defines both the event severity and their corresponding risk preferences. To measure this value, different functions $u(\alpha, \theta)$ will apply different penalties to the elementary score for individual combinations of (α, θ).

8.2.4 Eliciting risk preferences for decision-makers

The effects of infectious disease outbreaks are wide-ranging. Decisions supported by forecasts–such as whether to enact additional vector control strategies or increase ICU bed capacity–can have profound health, social, and economic impacts.

Forecasters and evaluators can account for these various factors via $u(\alpha, \theta)$, yet eliciting these utility functions for individual decision-makers is non-trivial. In the current work, we focus on how C/L ratios (α) could be defined for different event thresholds (θ), as a necessary first step for future input to custom functional forms for $u(\alpha, \theta)$. See[655] for a detailed exploration of how different proper scoring metrics apply different asymmetric penalties.

α and θ must be defined simultaneously and both require real-time updating. For example, suppose that a local health authority (the decision-maker) has determined that to protect against the loss of life from severe disease, it is necessary to increase/reallocate another 10 ICU beds to a district if more than 1,000 (θ) infections occur (the event) in a given week. ICU beds are made available at an economic cost if a chosen model (see below) forecasts infections exceeding the threshold in the next week with probability $\geq$ C/L (α). The economic costs of action (C) could depend not only on the additional beds themselves but also the probable length of patient stays, additional medical and staffing costs, and possibly indirect costs if beds are relocated from elsewhere. Meanwhile, if the event occurs without preventive action, preventable losses (L) could entail life costs from direct impacts of severe disease, co-morbidity, and macroeconomic impacts (e.g., greater workforce absenteeism by patients and relatives). Regarding life costs, the standard approach in the United States to quantify mortality risk reductions in monetary terms is the value of a statistical life, commonly estimated at about $10 million in USD2017[656], while in the United Kingdom the National Institute for Health and Care Excellence cost-effectiveness threshold for medical interventions is set to £30,000 per quality adjusted life year[657]. Quantitative estimates of costs for different actions and losses from epidemic impacts may also be available (e.g., for COVID-19[658,659,660] or dengue[661,662]) which can be used to calibrate C and L.

There will inevitably be residual uncertainty, ethical judgements, and legal implications. Many costs are unforeseen or difficult to quantify, yet forecast-informed public health decisions are high-stakes. In the absence of reliable estimates for all relevant quantities, forecasters and evaluators can perform sensitivity analyses and surveys (see Discussion and Section 8.4) to consider reasonable ranges of their values, while recognising that decision-makers may prioritise different costs depending on their preferences (e.g., prioritising impacts from loss of life by down-weighting possible contributions to C related to operational costs).

In general, C/L ratios for different epidemic event thresholds will vary i) by pathogen and its transmission history (e.g., immune landscape), and ii) by decision-maker and available resources (e.g., a higher C/L ratio in a resource-limited setting requiring greater certainty to minimise false alarms). These will also vary across space and time as epidemics and their impacts evolve alongside economic health. By evaluating the value of forecasts over a range of C/L ratios and event thresholds, determined by variation in costs arising both from uncertainty and differing decision-maker priorities and risk preferences, forecasters and evaluators can explore the Pareto frontier of forecasts that are performant simultaneously relative to multiple objectives. Finally, we emphasise that by evaluation only depending on the C/L ratio, it is only the relative cost of mitigation that is required for forecast evaluation.

8.3 Forecast value for informing epidemic decision-making

8.3.1 User-and-event-focused scoring metrics

When confronted with the introduction of a new pathogen, the coming season of an endemic pathogen, or an apparent rise or decline in an epidemic trajectory, decision-makers are faced with difficult and specific decisions (Figure 8.2 A), for which a forecasting model's *value* may vary. These decisions require forecasters and forecast evaluators to focus model development, model selection, and forecast evaluation on specific epidemic events, specific parts of the forecast distribution, and a specific decision-maker's risk preferences. Decisions are generally about actions drawn from a wider *action set*, and may be short-term interventions (e.g., public information campaigns) or longer-term plans (e.g., vaccination strategies).

Table 8.2: **Metrics of our framework:** Details of different metrics of forecast quality and value, including the information needed to compute metrics and relative advantages and disadvantages of each metric.

Metrics	Data/User Requirements	Advantages	Disadvantages
Relative Economic Value (REV)	Rough sense of C/L ratio and defined event threshold	Interpretability, quantifies the relative/operational value of model, and event-specific	Focus on fixed event, need for sense of C/L ratios, and does not use elementary scores
Brier Score (BS)	Defined event threshold (θ)	Event-specific, proper scoring rule, and can be decomposed [641,663] into meaningful statistical properties	Focus on fixed event and uniformly across all user risk preferences (i.e., all C/L ratios)
Pinball loss/Quantile Score	Fixed risk preferences (C/L ratio)	User-specific evaluation measure, consistent scoring function, and can also be decomposed into meaningful statistical properties	Does not measure whole predictive distribution and assumes fixed risk preferences across event thresholds
Murphy Diagram	Defined event threshold or user preferences	User- *or* event-specific, interpretable visualisations, and can trace out forecast value under uncertain event thresholds (θ) and/or different risk preferences (α)	More formal hypothesis testing may be required and crowded visualisations
Continuous Ranked Probability Score (CRPS) and	Sample or full probability distribution representation of forecasts	Proper scoring rule measuring whole predictive distribution, simplest measure of forecast value for uniform family of user risk preferences, and can be decomposed into interpretable components [647]	Neither event- nor user-specific and computationally expensive (for sample representation)
Weighted Interval Score (WIS)	Forecasts represented as quantiles	Proper scoring rule measuring the whole predictive distribution, converges to CRPS, and computationally cheaper to handle quantiles (instead of samples) of a predictive distribution	Neither event- nor user-specific and limitations (e.g., distribution tails) of quantile representation

Viewing each decision problem as a *binary* one (action or no action), the *value* of a model's probabilistic forecasts (e.g., of incident cases) is determined by its ability to inform the binary decision problem(s) for different decision-makers, given their corresponding utility functions $u(\alpha, \theta)$. Therefore, the *value* of a forecast depends on both the forecasting model and the decision-maker's utility function ($u(\alpha, \theta)$) for their policy question. This framing does not mean that a forecast is the sole basis for complex epidemic decision-making, but rather that the value of a forecast must be determined by its ability to inform a decision-maker.

Relative Economic Value

An approach in our framework to measure forecast value is Relative Economic Value (REV) [161,640]. For an event threshold (θ) and a decision-maker's corresponding C/L ratio (α), REV measures the value of a model's forecasts *relative* to a baseline/reference model's forecasts and a perfect model's forecasts. REV is a decision-theoretic framework often used in weather forecasting (e.g., [286,640,654]). It relies on knowing a plausible range of a decision-maker's C/L ratios (α) for the defined epidemic event threshold (θ).

For an event threshold θ, REV starts by converting the probabilistic forecast x_i into a classification of the binary event $y_i \in \{0, 1\}$ based on a given rule. Under the C/L framework, a forecasting model predicts occurrence of the binary event if the probability x_i exceeds the C/L ratio α. REV then uses the expected expense (E) that would be incurred by using the forecasting model's classifications for the binary event. This expected expense is calculated by combining the correct and incorrect classifications with corresponding costs and loss incurred, yet the calculation of REV relies only on knowing the ratio of costs and losses (see Appendix Section G.2):

$$\text{REV}(\alpha = C/L, \theta) := \frac{E_{\text{baseline}} - E_{\text{model}}}{E_{\text{baseline}} - E_{\text{perfect}}}. \tag{8.3.1}$$

Bounded between 0 (for the baseline model) and 1 (for the perfect model), REV is computed for different C/L ratios (α) while holding the event threshold (θ) constant. REV is then shown as a curve for different decision-maker risk preferences

(α), with higher values indicating superior relative performance for a given risk preference and event threshold θ (Figure 8.2 C). We use REV in our framework as an interpretable, event-focused measure of a model's operational/relative value over a baseline model for different risk preferences (Table 8.2).

User- or event-specific Murphy diagrams

A Murphy diagram[652] provides user- or event-specific measures of forecast value (Figure 8.2 C, Table 8.2), and extends beyond the REV by capturing the forecast value *and/or* overall predictive performance, using a proper scoring rule (BS) or consistent scoring function (pinball loss) without the need for a baseline model.

First, as a *user-specific* evaluation tool, the Murphy curve (Figure 8.2 C) visualises, for a fixed C/L ratio α, the mean elementary score (y-axis) for a given forecasting model against different epidemic event thresholds θ (x-axis) across *all* N test set observations:

$$\overline{\mathrm{S}_\alpha}(\theta) := \frac{1}{N}\sum_{i=1}^{N} S_\alpha(x_i, y_i) \tag{8.3.2}$$

where for observation i, $x_i \in \mathbb{R}$ is the model's forecast at quantile level $1-\alpha$, $y_i \in \mathbb{R}$ is the observed outcome, and S_α is the corresponding elementary score (eq. (8.2.6)) for the fixed C/L ratio α and a specific event threshold θ. In our framework, we include Murphy diagrams, a collection of Murphy curves that trace out, for a fixed decision-maker's risk aversion preferences α, how the value of different models' forecasts varies across different magnitudes of epidemic events (θ). Lower curves for a given model are preferable, and if curves for different models cross, there may be no model dominance[648,652]. The area under each Murphy curve is the mean pinball loss for that model. This visualisation improves on BS as it measures value for a specific decision-maker's risk preferences and different epidemic event thresholds, which may be advantageous when events are difficult to *a priori* define.

Alternatively, a Murphy curve can be an *event-specific* evaluation tool (Figure 8.2 C), holding the epidemic event threshold θ fixed and visualising a model's mean elementary score $\overline{\mathrm{S}_\theta}(\alpha)$ (y-axis) against different C/L ratios α (x-axis):

$$\overline{\mathrm{S}_\theta}(\alpha) := \frac{1}{N}\sum_{i=1}^{N} S_\theta(x_i, y_i), \tag{8.3.3}$$

where S_θ is the elementary score (eq. (8.2.6)) for the fixed event threshold θ and a specific C/L ratio α. The area under such a Murphy curve is the mean BS, and a Murphy diagram again visualises Murphy curves for different models. This diagram traces out, for a fixed epidemic event severity (θ), how the forecast value of different models varies with different risk preferences (α) for the decision-maker.

Caution is necessary, as: a user-specific Murphy diagram does not measure the whole predictive distribution and scores could be dominated by tail behaviour, while the event-specific Murphy diagram focuses on a single event threshold; neither diagram segments forecast performance into discrimination or calibration; varying the fixed event thresholds and/or decision-maker preferences is important; and formal inference and hypothesis tests may be required.

Other measures of forecast value

We briefly highlight other metrics that focus forecast evaluation on specific decision-makers and their priorities. Originating from weather forecasting[653], the diagonal score (Figure 8.2 B) assumes a fixed relationship between C/L α and the empirical probability of epidemic events $\pi(\theta)$ (i.e., how often different events occur). This makes it suitable for risk-intolerant decision-makers, as if the cost of action (C) remains fixed across event thresholds, rare extreme large event thresholds are associated with large losses (L).

Traditional classification tools include Receiver Operating Characteristic (ROC) curves and the Area Under the ROC Curve (AUC), as implemented in[46,223,574]. As ROC curves can fail to satisfy concavity requirements and AUC ignores

miscalibration[663,664], we recommend the use of DSC-MCB properties in the CORP-based decompositions (e.g., of BS) and/or concave ROC curves that use the recalibrated forecasts of the CORP decomposition[663].

8.4 Risk-averse epidemic decision-making and distributionally robust optimisation

Epidemic decision-making is high-stakes; decision-makers may prioritise risk-averse strategies that avoid the worst outcomes and remain robust to the considerable uncertainty in the dynamics of outbreaks. We described above uncertainty in a decision-maker's C/L ratios which could affect forecast value, model selection, and the metric-implied optimal decision. Meanwhile, observed data during an epidemic are imperfect, impacted by time- and space-varying issues such as reporting errors, delayed reporting, and underrepresentation. These issues, along with fast-evolving epidemic dynamics, can induce data distribution shifts and lead to metric-implied optimal decisions – from results for past test data, even if stratified by space and time (Section 8.2.2) – no longer being optimal when making the decision in real time on newly available data.

While the scoring metrics introduced above (Table 8.2) provide a general and flexible framework for evaluating forecasts, they are clearly limited and do not cover all relevant risk measures. We draw attention to distributionally robust optimisation (DRO) as an alternative or complementary setup for decision-making under uncertainty[665,666]. *Expected tail loss* (ETL) is a quantity central to the DRO literature (see[665,666] and Appendix Section G.7 for formulae and descriptions) that measures the risk associated with extreme events occurring at the upper tail of the loss distribution, defined as the average loss incurred among the worst α% of outcomes Y (e.g., hospitalisations). At many points during the COVID-19 pandemic, discussion and messaging surrounding public health measures implicitly invoked ETL as a focal risk measure. Emphasis on early lockdowns in China without complete knowledge of disease severity, for example, prioritised controlling tail risk. ETL and some other risk measures T of interest (e.g., the variance and mode of Y) are not *elicitable* from any proper scoring rule S, in the sense that there is no S for which predictions issued on the basis of T that uniquely minimize the score[431,667,668]. Therefore, evaluating forecasts based on S would yield conclusions inconsistent with the choice of T as a risk measure.

The distinction between scoring rules and risk measures discussed above points to the gap between the stochastic optimisation framework—within which scoring rules are situated—and DRO, respectively[666]. Here we draw attention to DRO as a body of work that may prove salient to the mitigation of future outbreaks. Although it has yet to take hold in the infectious disease literature, we view DRO as an important future tool to help address the limitations of our framework and the impacts of uncertain decision-making.

8.5 Predictability assessments for epidemic forecasting

Predictability analyses are another way to protect against the impacts of uncertainties, tail risks, and temporal dependencies (in epidemic data and forecast value). Unlike systems such as weather, predictability for an observed epidemic is not just a property of nature, as i) there can be a modifiable dynamical system with no governing laws, and ii) observed data are highly imperfect measures of the true nonlinear epidemic dynamics (e.g., incident infections). This raises the question of predictability. There is no universal definition of predictability, and we use a broad definition of predictability as the inherent randomness of the epidemic time series data (see below) that induces fundamental limits in the ability to predict the future values of the time series and therefore limits to forecast performance.

8.5.1 Introducing predictability for infectious disease forecasting

In our framework, we have described various metrics (see above and Table 8.2) to measure forecast quality and value. Forecast *quality* (or distributional accuracy) should be bounded above by the predictability of the epidemic data, yet how does the *value* of a forecasting model for informing decision-making relate to epidemic predictability? This is an open question that we begin exploring in the current work.

Arguably, a "good" forecasting model could be highly valuable for decision-making during epidemic stages of rapid change when there may be high unpredictability and a model could have *value* over a reference/naive model (e.g., seasonal average). The association between predictability and forecast value will depend on how a forecaster or evaluator defines both predictability and forecast value. Predictability is measured as a magnitude-free property of the underlying dynamical system. In contrast, forecast value is measured for an individual decision-maker and/or epidemic event, thereby depending on both the magnitude of the observed data and the set of models. Thus, as any measure of predictability is independent of the decision-maker, the models, and the magnitude of the data, a forecaster or evaluator should not expect predictability to have a one-to-one relationship with any measure of forecast value.

However, the forecaster, evaluator, and decision-maker need to know which outputs can be reliably reported and which forecasting models to use for a decision. Procedures are needed to evaluate how predictability has shaped forecast quality and value in the past and how to anticipate future epidemic predictability and its likely impacts on the utility of a forecasting model for informing decision-making. In our framework, we hypothesise that i) assessing past predictability of the epidemic time series will help to determine whether past model forecast performance was largely caused by the model and/or the (un)predictability of the epidemic, and ii) analysing past predictability can inform the likely future predictability and forecast performance of different models.

8.5.2 Permutation entropy

For estimating predictability, we follow [166] and use permutation entropy (PE), a model-free measure of temporal dependence in potentially nonlinear time series data. Entropy summarises uncertainty in a system. Shannon's entropy [591], for a random variable $X \in \mathcal{X}$, is defined as $H(X) = -\sum_{x\in\mathcal{X}} \mathbb{P}(X = x) \log \mathbb{P}(X = x)$.

For time series $\boldsymbol{X} = \{x_t : t = 1 \dots N\}$, PE works by first defining: i) the embedding dimension d which determines the segment length for measuring randomness/patterns in the data and ii) the time delay τ which determines the number of time steps between the compared measurements (see below). These are used to construct vectors: $\boldsymbol{x}^d_{t,\tau} = \left(x_t, x_{t+\tau}, \dots, x_{t+(d-1)\tau}\right)$. We fix $\tau = 1$ to focus on successive values of the time series.

Within each vector $\boldsymbol{x}^d_{t,\tau}$, entries of this segment of length d are then (totally) ordered by comparing their relative magnitudes. This produces an ordinal pattern π. Normalised PE is then defined by using the Shannon entropy on these patterns and normalising using the maximum entropy $\log(d!)$:

$$H_{d,\tau}(\boldsymbol{X}) = -\sum_{\pi} \frac{p_\pi \log(p_\pi)}{\log(d!)} \tag{8.5.1}$$

where p_π is the empirical probability of a specific ordinal pattern π, based on the frequency at which each pattern occurred.

To promote conservative estimates of PE, we again follow [166], and define predictability as:

$$\chi = 1 - min_{d,\tau=1}\left(H_{d,\tau}(\boldsymbol{X})\right), \tag{8.5.2}$$

where the minimisation searches over values from $d = 2$ to a user-defined maximum dimension (see [166] for more details). Using eq. (8.5.2), we can compare epidemic predictability (χ) across space and time using trailing windows of length N and region-specific time series data, noting that i) the window length N, over which PE is calculated, can be varied, but

Table 8.3: **Submitted model names and descriptions:** The following table includes our abbreviation, the original model name from each modelling team's submission, and the corresponding model description provided in [170].

Model name	Original name	Description
Ensemble	COVIDhub-4_week_ensemble	An unweighted average or median of submitted forecasts to the COVID-19 Forecast Hub
Baseline	COVIDhub-baseline	The median prediction at all future horizons is equal to the most recent observed incidence
CU-Select	CU-Select	A metapopulation county-level SEIR model for projecting future COVID-19 incidence cases and deaths
Karlen-pypm	Karlen-pypm	Finite time difference equations implemented as a general-purpose population modelling framework
CA-DELPHI	COVIDAnalytics-DELPHI	A SEIR model augmented with underdetection and interventions
RW-ESG	RobertWalraven-ESG	A multiple skewed Gaussian mathematical model fit

must be greater than $d!$, and ii) PE (and therefore χ) is global-magnitude-invariant as it compares relative magnitudes of observations. Values of χ vary between 0 (low predictability) and 1 (high predictability).

8.5.3 Embedding predictability in our framework

In predictability assessments of our framework (Figure 8.1), a forecaster or evaluator can compare the values of χ to results in trailing windows for the different metrics of Sections 8.2–8.3. Although low predictability *may* suggest low forecast quality, i) high predictability does not guarantee high forecast quality or value and ii) low predictability does not guarantee low forecast value. Understanding which models provide the most relative improvements in forecast quality or value over baseline/reference models during different regimes of predictability is one way to link predictability to forecast quality and value. Then, an estimated recent or future low predictability will have setting-specific implications – this may provide caution *or* optimism about using, for decision-making, a given model's forecast of the future.

Given that i) an operational forecast is often required to support decision-makers irrespective of how predictable an outbreak is, and ii) data distribution shifts are pervasive in epidemics, these predictability analyses after applications of metrics could be a useful sense-check and additional layer to model evaluation. These analyses could help quantify the practical or operational value at different stages of the epidemic and protect against future erosion of public or decision-maker trust. For example, values of χ could indicate regime changes, which could inform model selection or data collection as different models may have been more or less successful in the past under similar predictability regimes and/or further data may be required. Values of χ could also be used by forecasters as an input to a forecasting model or as an indicator of a future change in other epidemic indicators (e.g., in the time-varying reproduction number $R(t)$).

8.5.4 Extending predictability analyses

A key limitation of using eq. (8.5.2) is its univariate nature. An epidemic may be more predictable conditional on knowing other information, often provided by heterogeneous data streams. To quantify historical drivers of predictability, we propose analyses across space and time. For example, χ could be estimated alongside $R(t)$ to understand the association between predictability and stages of epidemic growth. By further associating results with those for the forecast metrics of the previous sections, these retrospective analyses can enhance our understanding of the reasons behind historical forecast performance, and this understanding can then be used for anticipating future predictability and future performance.

8.6 Applying metrics to epidemic forecasts

8.6.1 Application of our framework

We have introduced a conceptual framework for which the workflow (Figure 8.1) could be as follows. Given the available data, the first step for a forecaster or evaluator is to: i) co-define a relevant policy question with a local decision-maker, ii) identify a corresponding epidemic outcome to measure (e.g., incident cases or healthcare capacity), iii) produce or collect model-based probabilistic forecasts, iv) identify appropriate metrics and a decision-maker's preferences, vi) measure forecast performance for each model across statistical properties, over time, and across space, and vii) assess historical predictability of epidemic dynamics alongside past forecast performance to anticipate future reliability and model suitability, and viii) use metric-based results to choose the "optimal" model's forecasts for future decisions where optimality is only with respect to a given user and event. As decisions will not be made on the basis of a single question or epidemic outcome, this evaluation workflow should be repeated multiple times at each time point, using different policy questions and different outcomes and metrics targeting the same policy question.

We illustrate the framework with a brief application (Table SI G.2 for EPIFORGE 2020[564] checklist items), using state-level forecasts of weekly incident cases from the COVID-19 Forecast Hub (see [170,172,173] for details).

Our objective was not to repeat previous analyses, but rather to highlight i) how different metrics capture forecast value for informing simple policy questions, and ii) the implications of the statistical theory for different decision-makers and epidemic stages. There will be no one-size-fits-all approach for forecasting, evaluation, or decision-making, and any application should make clear the subjective choices made. Setting-specific considerations for the decision-maker include relevant policy questions, intervention action sets, costs and resources, and likely responses from the public. For the forecaster and evaluator, considerations for model development and selection include evaluation window sizes and spatial scales, both of which depend on practical data availability and requirements of public health authorities.

We studied the time period of 1 August 2020 – 15 January 2022 inclusive and restricted our focus to six models (Table 8.3) that submitted forecasts for all 73 weeks. We used the COVID-19 Forecast Hub baseline model as a reference model, which forecasts the most recently observed incidence as its median estimate with uncertainty based on historical differences between forecasted and observed values[172].

We were limited by the probabilistic forecasts being raw case counts and summarised using seven quantiles, preventing the use of recommended transformed scales for evaluation[213] and evaluation of the full forecast distribution (and all C/L ratios). We adopted a 12-week trailing window to examine predictability of the epidemic and time-varying performance of models. This enabled a realistic application, as a forecaster or evaluator would likely not want ensemble weights or models selected to change too rapidly, as there is already enough weight and model uncertainty in fast-evolving epidemics. To measure overall forecast quality, we used the WIS[638] alongside relative WIS (rWIS) – a scaled version of WIS that measures relative performance compared to the baseline model and accounts for potentially different numbers of forecasts for different time points and locations. We used the same weekly estimates as[170] for the instantaneous reproduction number, $R(t)$. All code and data used in our analysis are available at `https://github.com/cathalmills/forecast_evaluation`.

8.6.2 From metric to action I: Evaluation workflow results

We first evaluated the forecasts across the studied time period. Following our workflow (Figure 8.1), suppose a first policy question is the common task of predicting the future continuous value of weekly incident COVID-19 cases for each of the 51 studied states/territories. Using WIS and rWIS, for both forecast horizons of 1–4 weeks, the ensemble model generally outperformed other models, including across all time points (Figure SI G.1 A), for individual locations (Figure SI G.1 B), and in trailing windows of length 12 weeks (Figure SI G.1 C, SI G.3 – SI G.6). Outperformance over time and across space

was less clear for longer forecast horizons (Figure SI G.1 C, Figures SI G.5 – SI G.6). As WIS approximates CRPS for quantile-represented forecasts, the ensemble model generally produced forecasts with the most value for a *uniform* family of decision-makers (i.e., uniform risk preferences) when measured simultaneously across all epidemic event thresholds (see Section 8.2.3).

These forecasts have already been extensively analysed through the lens of WIS [170,433]. In epidemic settings, we are often interested in the many decisions for epidemic control faced by a specific decision-maker. A forecaster or evaluator can consider a risk-intolerant decision-maker with low C/L ratio (α) – action is inexpensive, relative to epidemic losses. In weather forecasting, a typical range for extreme events considered is α between 0.01 and 0.2 [286,669]. For both forecast horizons and a range of C/L ratios, according to the pinball loss and its CORP decomposition, the ensemble model produced simultaneously more discriminative and better calibrated forecasts (Figure 8.3 A). This indicates higher forecast value for the ensemble model for risk-intolerant decision-makers across the entire set of observed epidemic event thresholds. Such user-specific insights may be useful if the decision-maker could not, or did not wish to, define a specific epidemic event *a priori* (e.g., evolving dynamics), but had a fixed risk preference across event thresholds and over time. We further examined pinball loss for different decision-maker preferences over time (Figures SI G.15–SI G.16) and across space (Figures SI G.18–SI G.24), where the ensemble did not always outperform the baseline model for risk-intolerant decision-makers (e.g., Figure SI G.15 and SI G.18). However, the ensemble was the only model to provide a generally greater forecast value over the baseline model across different forecast horizons and most C/L ratios, aligning with the findings for WIS across the predictive distribution. We found greater relative value for more risk-tolerant decision-makers (higher C/L ratios) and for shorter forecast horizons.

Focusing on the axis of epidemic events (see Figure 8.2 B, right panel), we found that for classifying the upper 10% of each location's observed incident cases (location-specific event thresholds θ), the CORP-based decomposition of the BS indicated similar results – the ensemble providing more discriminative and better calibrated forecasts for decision-makers (Figure 8.3B). This event-specific analysis focuses on state-specific extreme events for a uniform family of decision-makers, yet ignores any likely risk preferences of the decision-maker that could expressed with a range of plausible C/L ratios, especially for extreme events.

User-specific (Figure 8.3 C.1) and event-specific (Figure 8.3 C.2) Murphy diagrams allow simultaneous assessment of forecast value for different epidemic events (or decision-makers) and a specific decision-maker (or a specific epidemic event). The Murphy diagrams for specific decision-makers (Figure 8.3 C.1, holding α fixed) trace a trend of the Karlen-pypm and ensemble model's forecasts often providing the most value for risk-intolerant decision-makers with low C/L ratios (for whom action is inexpensive, relative to the loss). We found less clear trends for rarer, larger event thresholds (θ) across forecast horizons. For more risk-tolerant decision-makers (i.e., higher C/L ratios), the ensemble model provided clearer value for longer forecast horizons (Figure SI G.26). These user-specific Murphy diagrams, while crowded and likely benefitting from further focus on event thresholds θ for specific policy questions, capture the value of the models' forecasts for informing a specific decision-maker's binary decision of whether to act based on the forecasted probability of the future (one or four weeks ahead) weekly incidence exceeding different event thresholds (θ). Focusing on location-specific extrema (location-specific event thresholds θ), the event-specific Murphy diagrams for BS (Figure 8.3 C.2) capture that the ensemble model provided the most value for classifying the upper 10% of each location's incident cases for most decision-makers across both forecast horizons (up to C/L ratios of 0.75). There is benefit in using both types of Murphy diagram, as the user- and event- specific visualisations can capture forecast value when there is uncertainty in the defined event (e.g., evolving dynamics) and/or decision-maker's C/L ratio (e.g., likely uncertainty about cost of intervention or epidemic event loss). This can inform decision-making under different uncertainties.

The REV does not capture forecast performance in terms of elementary scores, but provides an alternative, interpretable measure of relative value for different forecasting models over other simple, operational models. The choice of reference

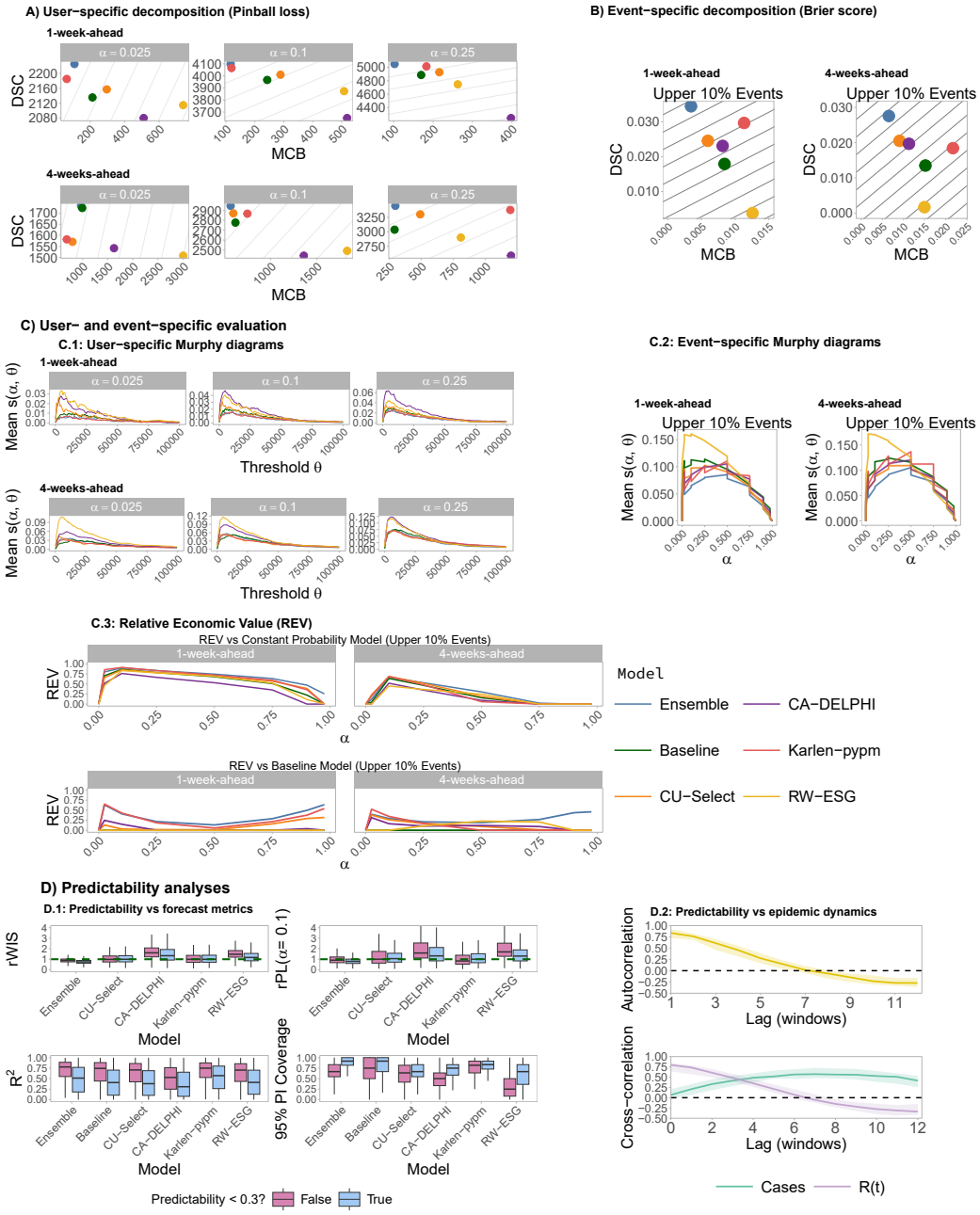


Figure 8.3: **COVID-19 forecast evaluation results** for application of our framework to state-level forecasts from 1 August 2020 – 15 January 2021, focusing on one-week-ahead and four-weeks-ahead forecasts. **A)** Pinball loss, a user-specific measure of forecast value, using discrimination (DSC) ability versus miscalibration (MCB). Higher DSC and lower MCB values indicate superior forecast value for a specific decision-maker's C/L ratio α. **B)** Analogous decomposition for the Brier Score, an event-specific measure of forecast value, focusing on the upper 10% of each location's observed cases. **C)** User- and event-specific evaluation measures of forecast value including event-specific Murphy diagrams (C.1) which capture the mean elementary score $s(\alpha, \theta)$ for a fixed C/L ratio (α) and varying event thresholds (θ), user-specific Murphy diagrams (C.2) which capture mean $s(\alpha, \theta)$ for a fixed event threshold (θ is the state-specific upper 10% threshold of cases) and varying C/L ratio (α), and Relative Economic (REV) diagrams (C.3) for the upper 10% of each state's cases. Lower values on Murphy curves and higher values on REV curves indicate superior performance of a model's forecasts for a given user-event combination. **D)** Predictability analyses include boxplots (D.1) of the relationship between predictability (binary thresholds here, further analyses in Tables SI G.3 – SI G.5, Figures SI G.27 – SI G.29) and forecast metrics, alongside median autocorrelation (D.2) of predictability and cross-correlation of predictability with cases and $R(t)$, time-varying reproduction number estimates. Shaded intervals capture the interquartile range across locations. Metric details are summarised in Table 8.1 and Figure 8.2.

model is crucial, and we used i) a probabilistic classifier that always predicts the binary event to occur with a fixed probability (similar to the climatology commonly used in weather forecasting[286,634]), and ii) the COVID-19 Forecast Hub baseline model[173].

Using the REV for the upper 10% of each location's incident cases (location-specific event thresholds θ), relative to a constant probabilistic classifier (forecasting occurrence of a binary event with probability 0.1, Figure 8.3 C.3 top row), we found that i) almost all models provided relative value for a range of decision-maker risk preferences, ii) REV was generally greater for shorter horizons, and iii) the ensemble model (alongside the Karlen-pypm model for one-week-ahead forecasts) provided the most clear relative value across the range of risk preferences. Although the relative value of models' forecasts was often lower when using the baseline model as reference, the ensemble model again provided the most clear relative value, including for risk-tolerant decision-makers (high C/L ratios). We repeated these REV analyses for different types of events, such as the upper 10% (15%) of all observed incident cases (Figures SI G.11–SI G.12) and for the upper 15% of each location's observed incident cases (Figures SI G.9–SI G.10 and Figures SI G.13–SI G.14), and found largely similar results (e.g., highest REV for these same two models and similar trends across C/L ratios and forecast horizons). In practice, the results from these REV diagrams could be filtered by a forecaster or evaluator to interpret them for a specific decision-maker's defined event, forecast horizon, and available action.

Relating forecast performance to predictability, higher predictability was moderately yet significantly (Table SI G.3) correlated with higher R^2 for each model's forecasts across different horizons (Figures 8.3 D.1 and SI G.27). Using relative measures such as rWIS (Figures 8.3 D.1, SI G.29) or relative pinball loss across different C/L ratios (Figures 8.3 D.1 and SI G.25), we estimated that several forecasting models generally had *greater* relative performance when epidemic dynamics were *less* predictable. However, greater relative performance for the given model during times of lower predictability, compared to times of high predictability, does not mean that the models necessarily provided improvements over a naive/baseline model at such times (Figure 8.3 D.1). It is necessary to understand their overall predictive performance during all regimes of predictability. We interpret the relative improvements alongside comparisons to the baseline (rWIS and relative pinball loss alongside predictability, Figures 8.3 D.1). We estimated generally consistent benefits, across space, time, and metrics, of using the ensemble model over the naive baseline model, for decision-making during less *and* more predictable times. The estimated relative improvements of the ensemble model, during high unpredictability, were often greater at longer forecast horizons (Figures SI G.15 – SI G.29).

These findings shed light on how predictability could have impacted the value of different models, yet can such predictability itself be anticipated and/or what drives different predictability regimes? Higher predictability within a window was strongly correlated with higher $R(t)$ estimates within the window (Pearson's[586] correlation $r = 0.70$, 95% CI = (0.68, 0.72)) and in past windows, while higher predictability was correlated with higher levels of incident cases in the recent past (Figures 8.3 D.2, SI G.30). This indicates that predictability was greatest during the epidemic growth and peak phases (as classified by[170]), and lower during the declining and nadir phases. Overall, we found generally low predictability, and this varied across space and time (Figures SI G.30 – SI G.31). These insights, while needing further exploration beyond our scope of evaluation, may be useful for informing model development and selection and for anticipating the likely future reliability of outputs. For example, lower $R(t)$ preceding lower predictability, or past performance during growth or declining phases, may be used to guide changes to a model or to an ensemble's weights. In future works, the recent and likely future predictability could be incorporated into the forecasting models or into the model selection procedures. For instance, we found that for the ensemble and several models, higher predictability was correlated with lower coverage of central 95% prediction intervals (PIs) (Figure 8.3 D.1, Table SI G.4), particularly for longer forecast horizons (Figure SI G.4). This reflects findings from[170] where coverage dropped during growth phases which we estimated to be generally more predictable. During such phases, central 95% PI coverage, even from the ensemble's forecasts, was often much lower than 95%, suggesting a need for updating of ensemble weights and/or models and their inputs (e.g., data

collection).

8.6.3 From metric to action II: Real-time decision-making

Above, we evaluated forecast quality and value *across* the 77 weeks. In practice, forecasters and evaluators will apply the framework for local decision-makers in specific locations at specific times. It is so far less clear how the forecasts could have informed decisions in real time for location-specific, important stages of the epidemic. Suppose now that a decision-maker in a given location is unknowingly at the onset of a location-specific large wave, which we define as the first date in 2021 when the location's median estimate of $R(t) > 1$. We focus on Arizona (a randomly chosen state) and describe pre-wave results in Appendix Section G.6.1.

Studying the COVID-19 wave, rWIS results for one-week-ahead forecasts remain comparable to pre-wave results (Appendix Section G.6.1). The same two models (Ensemble and Karlen-pypm) provided the best overall predictive performance, yet no model consistently outperformed the baseline for four-weeks-ahead forecasts (Figure 8.4 B). Focusing on specific events, co-defining *a priori* the policy question and epidemic event to measure before the wave is non-trivial and infinite choices exist. First, a decision-maker may be interested in whether future incident cases at a specified horizon breach a fixed extreme threshold, say $\theta = 20,000$ (an extreme value *before* the wave, Figure 8.4 A). The decision problem for this event could correspond to taking action based on the costs and preventable losses of maintaining a public health measure (e.g., additional ICU capacity, see Section 8.2.4) if incident cases $y \geq 20,000$. Using the CORP-based decomposition of the mean BS for this event (Figure 8.4 C, left), both the ensemble and CU-select model provided the most value across discrimination ability and calibration properties. However, if we add a decision-maker's risk preferences (Figure 8.4 C, middle), the event-specific Murphy diagrams capture how the optimal model varied by risk preferences. The analogous REV plots (Figure 8.4 C, right) are also crowded, yet illustrate differences in selected models across risk preferences and the low value for longer forecast horizons. In practice, we could provide the operational forecast value of each model for a specific decision-maker using a plausible range of their C/L ratios for this event threshold ($y \geq 20,000$).

On the other hand, an event threshold may become less meaningful if dynamics are evolving quickly and incident cases are predictably above/below this threshold. Varying this event threshold (θ), the user-specific Murphy diagram for pinball loss (Figure 8.4 D, right panel) captures how well the forecasts informed decisions for different event thresholds. These diagrams can also easily become crowded (e.g., due to a wide range of events on the x-axis, see Figure 8.4 D). Alternatives may be to focus on the CORP-based decompositions for BS or pinball loss (Figure 8.4 C–D) or to restrict x-axes of the Murphy diagram to event thresholds (θ) of interest.

Suppose a decision-maker cannot pre-define a fixed event threshold and/or will not base decisions on a single event threshold, yet has a fixed risk profile (C/L ratio α). Focusing on *all* observations of the wave for these fixed risk preferences, the pinball loss decompositions (Figures 8.4 D and SI G.35) capture how the ensemble provided the clearest value for risk-intolerant decision-makers. However, the chosen model for decision-makers varied according to their relative cost of mitigation (C/L ratio) and the forecast horizon. The models with the most value for a given decision-maker during the wave (Figure SI G.35 B) often differed to the models with the most value before (Figure SI G.35 A) the wave. We found similar differences in models selected before versus during the wave when using the Brier score, REV, and relative pinball loss (Figure SI G.36–SI G.37). Collectively, these results suggest that for policy-focused model development and representative evaluation, forecasters and evaluators need to i) zoom in on user- and event-specific measures and ii) preserve the spatiotemporal structures of epidemics in evaluation and model selection. Otherwise, averaging over these heterogeneities may result in inaccurate measures of forecast value and a failure to provide different decision-makers with the optimal model's forecasts.

Retrospectively, we can analyse how well the peak values were predicted, for example, using BS or REV for the upper 10% of incident cases during the wave. This assessment of wave-specific extremes can be used by forecasters and/or

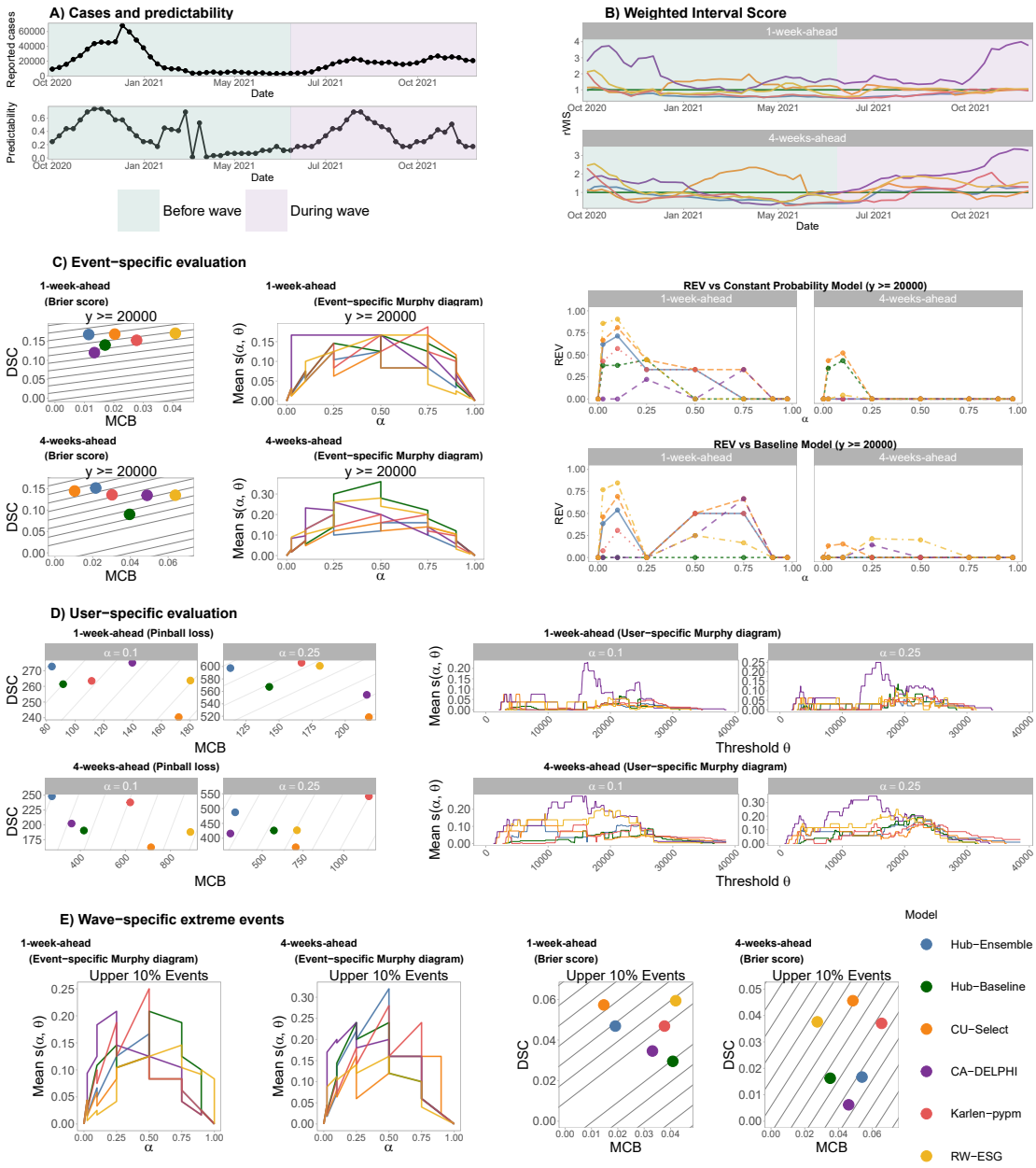


Figure 8.4: **Evaluation for real-time decision-making during a COVID-19 wave in Arizona: A)** Reported incident cases and estimated predictability before and during a COVID-19 wave that started (based on median $R(t) > 1$) in July 2021). **B)** 12-week trailing window relative Weighted Interval Score (rWIS) for all models across two forecast horizons before and during the wave. **C)** For two forecast horizons, Brier Score decomposition and event-specific Murphy diagrams, alongside a Relative Economic Value diagram, where the event is pre-defined as incident cases $y \geq 20{,}000$ (based on values before the wave). **D)** For two forecast horizons, Brier Score decomposition and event-specific Murphy diagrams for risk-intolerant decision-makers. We provide further visualisations for other risk-intolerant decision-maker preferences and risk-tolerant decision-makers in Figures SI G.34 – SI G.37. **E)** Event-specific Murphy diagrams and Brier Score decomposition and event-specific Murphy diagrams for forecasting the upper 10% of incident cases during this COVID-19 wave. The studied period for the wave ends on January 15 2022.

evaluators to build better models going forward and/or to assign ensemble weights. Here, the best models across the C/L ratios for predicting this event again varied by different forecast horizons (Figure 8.4 E, right) and by different decision-makers' preferences (Figure 8.4 E, left), while the best models differed to those chosen for the upper 10% of events before the wave (Figure SI G.32).

Compared to the pre-wave results (Appendix Section G.6.1), similar predictability trends persisted during the wave (Figures 8.4 A and SI G.38–SI G.39), as predictability rose and fell alongside the epidemic trajectory – $R(t)$ was strongly correlated with estimated predictability (r = 0.80, 95% CI: (0.62, 0.91)) during the wave. Low predictability was again correlated with lower pointwise accuracy for models, and during such low predictability regimes, some models had greater value relative to the baseline model for the uniform family of decision-makers, but not necessarily individual decision-makers (Figures SI G.32 –SI G.37 and Figure SI G.39). Therefore, given that sharp predictability changes were also estimated towards the start and at the peak of the wave, it is perhaps unsurprising that the best models varied (Figures SI G.32 – SI G.37) i) for different decision-makers, ii) for different events, iii) over time within the wave, and iv) compared to the pre-wave period. Looking forward, these insights (e.g., predictability having consistent impacts on forecast performance), which benefit from our hindsight, suggest the need for model development and selection, ensemble weighting, and data collection to condition on predictability, in addition to time, space, and the decision-maker's preferences and priorities.

8.7 Discussion

Forecasting epidemics is difficult. The future that is predicted is both nonlinear and highly stochastic[99], and inherently uncertain[103,470]. Working towards more timely and reliable predictions, our understanding of infectious disease forecasting has developed recently. Methods for evaluating forecast performance have evolved substantially, yet it was unclear how this statistical rigour translates into measuring the forecast's value for informing a decision-maker's dilemma; to act or not. A disconnect between forecasters and decision-makers and a lack of interpretable and operationally relevant forecast outputs have long been challenges in the operational use of forecasts to support decision-making[3]. Here, for forecasters aiming to build better models for decision-makers' needs and for forecast evaluators aiming to evaluate how well the models meet these needs, we provide a framework that contributes procedures to: i) recast forecasting models' statistical strengths and weaknesses into policy-relevant statistical properties across space and time, ii) select the "optimal" model for a decision-maker's individual policy question and risk preferences, and iii) link the limits of predictability for the epidemic to model selection and forecast value. The aim is that this user-focused framework provides a foundation for improving forecasting models and measuring the value of their forecasts by their ability to inform the uncertain decision-making of infectious disease outbreaks.

As shown in our applications, our new framework does not replace best practices in forecast reporting[564], evaluation techniques[170,433,473,670], forecast model development (e.g., time series cross-validation[671,672]) or standard diagnostics (e.g., calibration and reliability assessments[648,673]).

Our framework establishes connections between many best practices and extends to other branches of science (e.g., weather forecasting and decision theory). Importantly, the user- and decision-focused conceptual framework presents evidence and does not provide advice (which would involve subjective policy guidance)[18,674]. The framework allows the forecaster and evaluator to provide a decision-maker with the forecasts of the "optimal" model for their decision problem, based on past performance for the set of models and the defined policy question, action set, and risk preferences. The optimality conditions are highly specific and must be pre-defined, yet our applications suggest how to incorporate uncertainties and misspecification in decision-makers' risk preferences and priorities. From a decision and social theory perspective, our user-focused prioritisations reflect theory that a group-level utility function which fairly weights individual preferences is mathematically problematic and arguably unattainable[675,676]. Given that individuals have different risk preferences, there is no single scoring function for the group that allows all individual users to select the same forecasting model for the given task. Instead, we incorporate a decision-maker's individual policy question and risk preferences to determine the model(s) that have provided the most value for informing past decisions. Similarly to[160], we argue that this

decision-theoretic framework enhances the credibility and defensibility of uncertain policymaking. Iterative application of our evaluation framework for different policy questions connects with other research focused on decision problems [3,160] as part of *integrated* assessments for scientific advisors and policymakers. Decision-making before/during epidemics will never be based on a single policy question or forecast, epidemic event/outcome, or branch of science, yet if forecasting is to take place, evaluation must measure how well the forecasts inform decisions.

For operationalisation, collaboration with decision-makers will be essential to improve our proposed workflow (Figure 8.1). Models, evaluation metrics, and epidemic trajectories can be complex, and stakeholders have different priorities and backgrounds. Communicating uncertainty to both technical experts and decision-makers is essential yet non-trivial [18,203], challenged by i) a lack of agreement about uncertainty and the language of probability, and ii) the many nuances of an epidemic forecast (see below) [15,158]. Clear, timely, and succinct communication will involve decreasing levels of technicality as information is passed up the decision-making chain/process, while preserving key modelling outputs and uncertainty. Similarly, how a scientist represents data (e.g., binary versus continuous [677]) and communicates uncertainty [678] affects how a user interprets the strength of evidence. Future behavioural studies could quantify the optimal ways to communicate evaluation and forecast results to different decision-makers. This could involve traffic light schemes and warning impact matrices [679], or interactive dashboards [540,670]. All of this will benefit from continuous dialogue between modellers, social scientists, and decision-makers [18,203].

Recent guidelines for prediction tools advocate for a 3-U (useful, usable, and used) framework [280], which we embrace in the context of evaluation with public-health-relevant quantities, user-focused evaluations and visualisations, and metric-based decision-making. For example, the C/L model appeals to decision-maker preferences around the world for possibly crude but interpretable thresholds [205]. Future post-event surveys and analyses should investigate how well the framework addressed the specific needs of decision-makers for informing decisions, and which metrics to focus on to provide reliable *and* timely information.

A key next step will be to elicit estimates of decision-makers' C/L ratios (α) in real time for different event thresholds (θ), requiring careful setting-specific considerations (Section 8.2.4). Without plausible ranges for these C/L ratios, the framework could remain abstract. Eliciting preferences will benefit from i) regular consultation and surveys with decision-makers and experts across multiple disciplines, and ii) economic evaluation literature for different user perspectives [680]. Real-time surveys could define the distributions of users' C/L ratios and event thresholds for different epidemic questions. For example, [681] developed a survey-based user BS to understand how forecast users combine objective probabilities with their subjective feelings when making decisions based on weather forecasts. Similarly, additional post-event analyses could help to better determine users' C/L ratios and policy priorities by studying their responses and actions, and the impacts on epidemic dynamics. There will always be both heterogeneity and uncertainty in such C/L ratios as decisions before and during an epidemic are multidisciplinary tasks with ethical and human components [160]. Given the feedback loops, transience and uncertainty in C/L ratios, reinforcement learning and/or model-predictive control may also be useful for iterative adjustments, alongside recent frameworks for evaluating cost-effectiveness of different interventions [659,682] and for integrating public health ethics in models [683].

The C/L model is grounded in decision theory, yet remains a simple abstraction of reality, and future work may consider asymmetric utility functions, three-way decision problems, continuous action sets and stochastic policies, and/or multi-objective algorithms for decision-making using different policy questions and metrics. Our framework components could also be embedded in recent multi-model approaches for policy decisions under uncertainty [105,443].

Depending on the target quantity, the predicted future can be modifiable, modifications that a forecasting model generally does not anticipate. The "best" model may be the one that promotes the decision-maker to implement interventions that have direct impacts on epidemic dynamics, which can differ from the model that produces the most accurate forecast of the realised future. Thus, the actions of a decision-maker can impact the very probability of the epidemic events that a model

aims to forecast. A probabilistic forecast integrates over uncertainty in many different counterfactual scenarios and is a special, *unconditional* case of a scenario projection [158] – a model's best estimate of what *will* happen.

Forecasts may be operationally advantageous for empowering decision-makers to make decisions based on predicted probabilities for the most likely future states of the world. This avoids the difficulties of probability statements and evaluation in inherently counterfactual scenario projections. However, forecasts can suffer from irreducible error due to unknown and possibly model-based modifications in the intervening period after the forecast date. This can yield limits to predictability, and such feedback from the future means that forecast evaluation, without adjustment, can measure both the adequacy of the forecasts and the effectiveness of policy responses [559]. These issues are exacerbated (alleviated) for longer- (shorter-) term forecasts where model-based future policy changes could be more (less) likely. Two potential ways to address these feedback loops include: i) directly incorporating all different policies with corresponding probabilities in the integration over all possible future states of the world, highlighting the need for modeller and decision-maker dialogue, or ii) using scenario projections for these quantities, which can consider the "what-if" situations of different future policy decisions [158]. We suspect that scenario projections, possibly paired with our evaluation framework, may be more suitable when there is greater uncertainty about future policy changes (e.g., longer-term horizon seasonal predictions), while unconditional forecasts may be more suited to short-term horizons. We highlight here that this feedback loop is not always an issue for forecasting as epidemiological quantities can be largely invariant to interventions (e.g., incident cases after adding to ICU capacity). All of this highlights the need for close collaboration with decision-makers to maintain clarity about the future policies, the future that a forecast/scenario projection is targeting, and the associated utility, limitations, and implications of the prediction model.

Forecasting and scenario projections remain two strands of the scientific toolkit to inform decision-makers. Epidemics are not simply a prediction problem. A reliable prediction does not necessarily convert to effective action. Without understanding *why* or *how* transmission has occurred, predictions cannot inform targeted strategies to control an epidemic. Our framework should be used alongside inference of other key epidemiological quantities, variant/serotype predictions, nowcasts of the present dynamics, analyses of intervention effectiveness, and many other tasks [15]. Our evaluation framework could be adapted for such tasks, for example for outbreak risk (is $R(t) > 1$?). Beyond modelling, experts from other fields (e.g., health economics and social theory) should be embedded in multidisciplinary decision-making processes for an infectious disease outbreak [160].

The applications of our framework have several limitations. We evaluated forecasts on the imperfect reported COVID-19 case data without any data on policy decisions and their impacts or time-varying and/or space-varying reporting rates. These influences may vary within a studied population and for different infectious diseases. Our applications therefore reflect the realities of data collection, modelling, and human responses in real-time outbreaks. However, our predictability metric (PE) is robust to both observational and dynamical noise, while we focus on stable performance over 12-week windows and across space to limit sensitivity to surveillance artefacts. Epidemiological delays are also commonplace, and delayed case reporting impacts the evaluation process and the communication of timely information for decision-making [165,631]. This suggests using new methods for estimating delay distributions [113,114,157] and for nowcasting [540,573,684]. In the future, forecasters and evaluators will benefit from i) extensions to DRO (Section 8.4), ii) further diagnostic checks (e.g., reliability diagrams [663]) iii) formal statistical testing [685,686,687], and iv) forecast post-processing techniques [426]. Finally, it will be instructive to analyse the relationship between evaluation metrics and different model types, to inform both model and ensemble development.

Our predictability analyses are a stepping stone towards more concrete links between predictability and forecasting. Using historical predictability to anticipate future predictability is imperfect. Historically predictable epidemic dynamics may no longer be predictable during fast-evolving epidemics (e.g., new SARS-CoV-2 variants). We could address this with i) predictability assessments that are conditional on covariates, ii) regular collection of heterogeneous data to anticipate

such changes, and iii) extensions to multivariate PE[688]. PE is also only one way to estimate predictability, and alternatives include permutation dependence[689] or learning and irreducible errors[690].

In short, our new framework shifts the perspective of forecast evaluation from the forecaster to the decision-maker. Recognising that there is no one-size-fits-all approach to forecasting, evaluation, or decision-making, we provide principled procedures for forecasters and evaluators to measure and improve the value of forecasting models for informing and empowering decision-makers in infectious disease outbreaks.

Chapter 9

Other publications

In addition to the manuscripts described above, I have been involved in further publications as an author/co-author. I include the abstracts for two of these manuscripts below.

9.1 Artificial Intelligence for Infectious Disease Epidemiology

Moritz U. G. Kraemer, Joseph L.-H. Tsui, Serina Y. Chang, Spyros Lytras, Mark P. Khurana, Samantha Vanderslott, Sumali Bajaj, Neil Scheidwasser, Jacob Liam Curran-Sebastian, Elizaveta Semenova, Mengyan Zhang, H. Juliette T. Unwin, Oliver J. Watson, Cathal Mills, Abhishek Dasgupta, Luca Ferretti, Samuel V. Scarpino, Etien Koua, Oliver Morgan, Houriiyah Tegally, Ulrich Paquet, Loukas Moutsianas, Christophe Fraser, Neil M. Ferguson, Eric J. Topol, David A. Duchêne, Tanja Stadler, Patricia Kingori, Michael J. Parker, Francesca Dominici, Nigel Shadbolt, Marc A. Suchard, Oliver Ratmann, Seth Flaxman, Edward C. Holmes, Manuel Gomez-Rodriguez, Bernhard Schölkopf, Christl A. Donnelly, Oliver G. Pybus, Simon Cauchemez, and Samir Bhatt. Artificial intelligence for modelling infectious disease epidemics. *Nature*, 638(8051):623–635, February 2025. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-08564-w. URL https://www.nature.com/articles/s41586-024-08564-w

The perspective piece entitled "Artificial Intelligence for Infectious Disease Epidemiology" was published in *Nature* in February 2025[53]. My role in this manuscript was to contribute to the discussion of i) the uses of AI across epidemiological modelling tasks such as statistical inference, forecasting, and prediction, and ii) the integration of AI with modern/novel data sources and different modelling methods.

Abstract

Infectious disease threats to individual and public health are numerous, varied, and frequently unexpected. Artificial intelligence (AI) and related technologies, which are already supporting human decision making in economics, medicine, and social science, have the potential to transform the scope and power of infectious disease epidemiology. Here we consider the application to infectious disease modelling of AI systems that combine machine learning, computational statistics, information retrieval, and data science. We first outline how recent advances in AI can accelerate breakthroughs in answering key epidemiological questions and we discuss specific AI methods that can be applied to routinely collected infectious disease surveillance data. Second, we elaborate on the social context of AI for infectious disease epidemiology, including issues such as explainability, safety, accountability, and ethics. Finally, we summarise some limitations of AI

applications in this field and provide recommendations for how infectious disease epidemiology can harness most effectively current and future developments in AI.

9.2 Systematic review of the effectiveness of social distancing measures and lockdowns during the COVID-19 pandemic

Caitriona Murphy, Wey Wen Lim, Cathal Mills, Jessica Y. Wong, Dongxuan Chen, Yanmy Xie, Mingwei Li, Susan Gould, Hualei Xin, Justin K. Cheung, Samir Bhatt, Benjamin J. Cowling, and Christl A. Donnelly. Effectiveness of social distancing measures and lockdowns for reducing transmission of COVID-19 in non-healthcare, community-based settings. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381(2257):20230132, October 2023. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2023.0132. URL `https://royalsocietypublishing.org/doi/10.1098/rsta.2023.0132`

This systematic review was commissioned by the Royal Society and published as part of the theme issue "The effectiveness of non-pharmaceutical interventions on the COVID-19 pandemic: the evidence". My role was as co-author and reviewer.

Abstract

Social distancing measures (SDMs) are community-level interventions that aim to reduce person-to-person contacts in the community. SDMs were a major part of the responses first to contain, then to mitigate, the spread of SARS-CoV-2 in the community. Common SDMs included limiting the size of gatherings, closing schools and/or workplaces, implementing work-from-home arrangements, or more stringent restrictions such as lockdowns. This systematic review summarised the evidence for the effectiveness of nine SDMs. Almost all the studies included were observational in nature, which meant that there were intrinsic risks of bias that could have been avoided were conditions randomly assigned to study participants. There were no instances where only one form of SDM had been in place in a particular setting during the study period, making it challenging to estimate the separate effect of each intervention. The more stringent SDMs such as stay-at-home orders, restrictions on mass gatherings and closures were estimated to be most effective at reducing SARS-CoV-2 transmission. Most studies included in this review suggested that combinations of SDMs successfully slowed or even stopped SARS-CoV-2 transmission in the community. However, individual effects and optimal combinations of interventions, as well as the optimal timing for particular measures, require further investigation.

Chapter 10

Discussion and conclusion

Each manuscript has its own discussion where we describe contributions, limitations, and future extensions. Instead, this chapter provides a detailed objective- and theme-structured discussion (Figure 10.1) of how the research has addressed the stated objectives of the thesis (Table 1.1 and Figure 1.1). Within each objective and research theme, we will also reflect on our methods and analyses in a wider context, the strengths and limitations in different settings, and the implications for future research and policy directions. We will close this thesis with our concluding thoughts.

As we will discuss, our modelling theory and analyses for infectious disease threats have allowed us to reach the thesis' research and public health goals: i) learn from the past, ii) monitor the present, and iii) prepare for the future.

10.1 Learning from the past: Relative roles of epidemic drivers

Understanding (quantitatively) *why* and *how* transmission of an infectious pathogen occurred is vital, as without it, epidemic control is likely impossible. It is through such an understanding of epidemics that a researcher can develop and weigh quantitative evidence to inform decision-making and epidemic control strategies[3,99,435]. We aimed to provide this understanding in this thesis. Adopting and extending methods and lessons from the COVID-19 pandemic and other outbreaks, we developed multi-model approaches that intersect many branches of science and incorporate heterogeneous data streams. We used these methods to probabilistically quantify the relative roles of different human and environmental drivers of dengue epidemics across space and time. Applying our exploratory and predictive approaches at varying spatial resolutions, we unified and extended existing approaches to model-based inference of dengue epidemics in resource-limited settings.

10.1.1 Coalescing human, biological, and environmental drivers of dengue

In an era in which climate change and its interactions with human and biological modulators (Figure 2.1), continue to increase the incidence and distribution of climate-sensitive infectious diseases, quantifying the relative role of climatic drivers in epidemic dynamics is crucial for informed public health policy[261,351,352,691,692]. Similarly, public disagreement and conspiracies persist about the climate crisis and its impacts[693], as well as about infectious diseases themselves[694,695]. It is therefore important that researchers provide and communicate justifiable scientific evidence on the roles of climate and human behaviour in shaping human health[696,697].

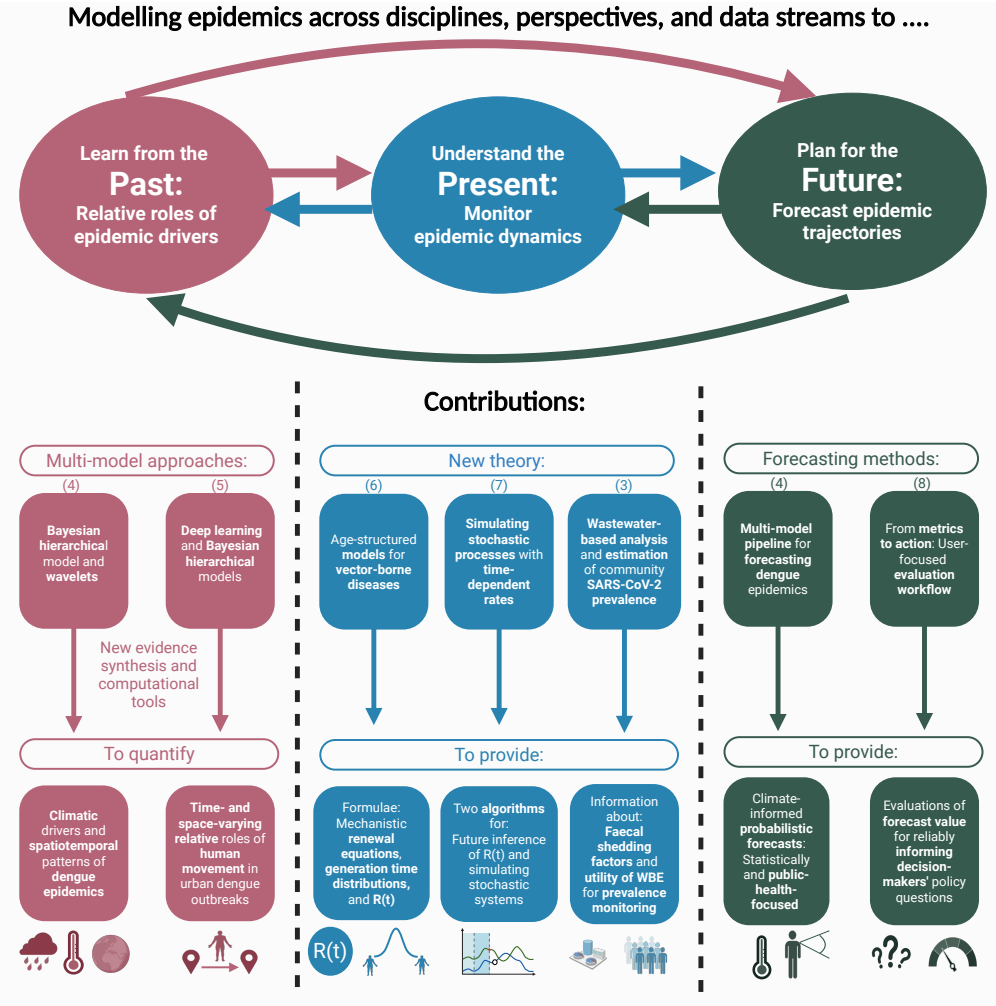


Figure 10.1: **Thematic overview of contributions and discussion**. In this thesis (Table 1.1 and Figure 1.1), we aimed to develop modelling theory and analyses to i) learn from the past, ii) understand the present, and iii) plan for the future. Here, we visually summarise our contributions towards each of these objectives, where number labels indicate the relevant chapter. The contributions are separated into the theoretical advances which then enable practical, public-health-focused outputs. In each section of this discussion chapter, we will discuss and visualise (Figures 10.2 – 10.4) each of these contributions further, alongside strengths, limitations, and future directions for each strand of work.

Epidemic drivers and dynamics across space and time

In a pair of analyses, we researched the dynamics of dengue epidemics over a twelve-year period in northern Peru – an area that has been at the epicentre of recent dengue-induced public health crises. In Section 4.1, at a coarse spatial resolution, we developed Bayesian hierarchical models to characterise the risk profile of the delayed, nonlinear effects of climatic and weather conditions on incident human dengue cases in three departments of Piura, Tumbes, and Lambayeque. In these models, we further modelled department-specific non-climatic seasonality, spatial autocorrelation, year-to-year heterogeneity, and epidemic momentum. The risk profiles that we estimated for climatic influences on dengue incidence were plausible and actionable, in terms of mosquito biology and ecology and local human modulators. We found i) a consistently elevated (diminished) risk for higher (lower) past temperatures throughout the preceding four months, ii) an elevated risk for higher recent rainfall and sustained drought conditions over longer durations between 1–4 months, and iii) an elevated risk for extreme values of the ICEN ocean sea surface temperature index (a Peru-specific El Niño indicator) at long durations of 2 – 4 months. These results for delayed climatic influences are intuitive given the mosquito life cycle and

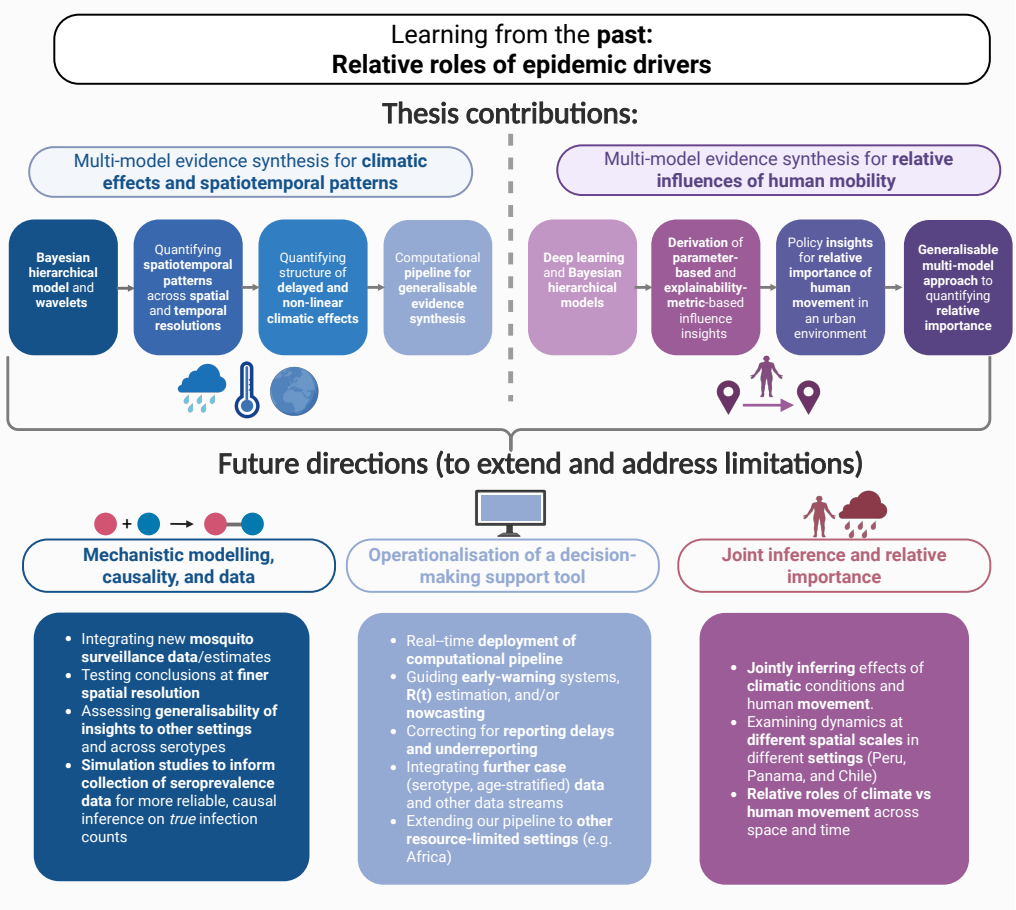


Figure 10.2: **Learning from the past:** In Chapters 4 – 5, we introduced methods and analyses to quantify the relative roles of dengue epidemic drivers in different settings at different spatial scales. We summarise these theoretical and practical contributions (in further detail than the left column of complementary Figure 10.1), alongside the future directions to extend our methods and address limitations.

the DENV transmission cycle (Chapter 2), as well as the time from human infection to self-reporting and any case-reporting delays.

Putting these climate-incidence results in a wider context, we do not expect identical risk profiles in other settings with different human and mosquito populations. Like previous studies, we cannot causally control for all population, data, or biological heterogeneities within and between nations. Different immunity levels, reporting rates, and serotypes are just some examples of these heterogeneities, and these can vary across space and time within studied regions and studied periods. Comparison of DENV and other arboviruses between settings is often cautioned for such reasons, including in WHO arbovirus surveillance reports[698]. Nevertheless, similar results have been reported for the delayed and nonlinear effects of minimum temperature in neighbouring Brazil[78]. The cumulative risk profiles for sustained temperature exposure also largely agreed with laboratory results for optimal temperatures in DENV transmission[82]. Likewise, researchers have estimated comparable lagged effects for precipitation and drought indices in Barbados[168], China[525], and Mexico[524], while we unveil the ICEN risk profile for Peru for the first time.

In the follow-up manuscript of Section 4.2, we obtained largely similar results when we modelled dynamics in finer spatial resolutions at a province level, with the exception of the ICEN. The estimated complex and less intuitive risk profile for the ICEN may be due to uncaptured causal modulators (see below, Figure 2.1, and Chapter 4) and/or difficulties caused by data at finer spatial resolution (e.g., case reporting).

In each analysis of Chapter 4, we strengthened the robustness of conclusions by using i) cross-validation techniques

such as leave-one-out and leave-future-out/time series cross-validation, ii) information criteria that measure out-of-sample generalisability, and iii) sensitivity analyses (see below). In the follow-up manuscript of Section 4.2, we complemented the Bayesian hierarchical model with wavelet methods to study the drivers and dynamics of short- and long-term cycles of dengue epidemics in different provinces. We further expanded our evidence synthesis by working with researchers and public health authorities who provided more detailed knowledge of local epidemic drivers and dynamics. In doing so, we unveiled spatiotemporal patterns and space-varying climatic and human drivers. These included highly synchronous cycles driven by pronounced climatic forcing in the more northerly and more westerly/coastal provinces. We estimated that high spatial synchrony across a twelve-year period was further influenced by shorter pairwise distances between provinces, and although this may indicate the importance of human movement for synchronising dengue epidemics, we did not explicitly model human movement due to data constraints.

Climatic drivers primarily influence human dengue epidemics indirectly via mosquito biology and ecology, yet humans remain crucial for vector-borne transmission. The mechanisms behind human mobility as drivers of dengue epidemics are less well understood versus climatic effects. However, our understanding is improving. More recently, the roles of short- and long-range human movements have been established in the seeding events and continued spread of dengue outbreaks[85,86,195,235,575,576,577]. Despite these advances, the relative importance of human mobility in driving dengue epidemics at different stages for different connected geographies is less well understood. Understanding *when* and *where* human mobility is important for cost-effective epidemic control – informing the optimal timing and targeting of public health interventions in resource-limited settings. We approached this knowledge gap by developing i) a new multi-model approach that again exploits the benefits of different modelling disciplines – mobility-informed Bayesian hierarchical and deep learning models, and ii) an evidence synthesis approach that blends parameter inference and information criteria with out-of-sample probabilistic predictions and forecast scoring metrics (Chapter 5, Figure 10.2). We focused on dengue outbreaks in neighbourhoods of an urban, endemic setting of Fortaleza, Brazil, and fitted our models using weekly incident dengue case data, population data and bus transportation data. To proxy weekly human movement flows between neighbourhoods, our models relied on bus transportation data. These data have been validated as representative in previous work and have had many uses beyond epidemic modelling[581,582,583]. Although we designed models for available neighbourhood-level data in Fortaleza, the models themselves are generalisable, computationally cheap, and conceptually simple. The influence of human mobility enters simply as a covariate. This approach has the benefit of computational efficiency, but at the expense of inferring the entire network structure. We argue that the bigger theoretical advances lie in our evidence synthesis approach.

To quantify the relative importance of human mobility in urban dengue outbreaks, after developing the models in a separate model development period, we i) derived new formulae for our Bayesian hierarchical model which estimate relative importance using learned parametric relationships, and ii) developed analogous quantities for our deep learning model that build on the popular explainability framework of SHAP values[418]. We validated our models by measuring out-of-sample generalisability using information criteria and scoring rules. A strength of this new evidence synthesis technique is that we avoided the dependence on a single model for public health policy insights. Instead, we introduced a multi-model approach that stretches across disciplines, emphasises statistical rigour via probabilistic validation, and leads to several representative models. This approach reduces the sensitivity of conclusions to the structural assumptions of any given model and reflects the consensus multi-model approaches that were often recommended during the COVID-19 pandemic[18,103,443].

We identified human movement as a constant important driver of the urban dengue dynamics and of consistently greater importance for driving dengue outbreaks relative to within-neighbourhood transmission. These results were somewhat surprising, as it was possible that human mobility could be most important in the early stages of an outbreak and thereafter influence transmission in the city to a lesser degree. To understand why such dynamics might have occurred and whether our multi-model approach gave plausible insights, we performed clustering- and entropy-based mobility analyses. These

analyses, along with insights from local researchers and data collectors, allowed us to further contextualise our results. We suspect that our importance findings could be a consequence of i) an almost homogeneous presence of mosquitoes throughout the city and over time, and ii) a high mixing between neighbourhoods with minimal mobility clusters for which mobility sustains the longer outbreaks observed. These align (respectively) with findings of mosquito adaptation from the limited surveillance data in the city[579] and with the results for COVID-19 outbreaks in well-mixed cities, relative to hierarchical cities[336]. We also estimated that the relative importance of individual neighbourhoods to onward transmission in other neighbourhoods was relatively constant within and across dengue seasons, in part due to relatively stable commuting patterns. It would be worth exploring how these results generalise to other settings such as a city without such a long history of transmission or other spatial scales. Similarly, it is unclear how these results could differ in Fortaleza under different commuting patterns, or by the inclusion of biological influences and other human mobility data streams such as long-distance movements.

Our new methods were designed for questions about the relative importance of human mobility, yet the multi-model way of thinking, model development procedures, and derived formulae could be easily extended to integrate further heterogeneous data sources and/or to study other drivers of dengue epidemics or other dynamical systems. For example, suppose we have a new outbreak of a climate-sensitive infectious disease and we want to know the relative roles of a climatic driver versus human movement. Further, assume that we can develop a well-validated Bayesian hierarchical model which jointly infers the effects of climate and human movement on incidence rates in a region. This model could combine the DLNMs of Chapter 4 for climatic effects with the mobility-weighted incidence parameterisation of Chapter 5. Taking the ratio of the two estimated effects for the competing processes provides an estimate of their relative importance. By fitting the model in rolling windows and using standardisation in each window, a researcher could quantify how the relative role of climate versus human movement varies between regions and within outbreaks. Likewise, conditional on having a deep learning predictive model of incidence rates in a region, a similar procedure could be repeated, involving our extended explainability metric (*r*SHAP, see Chapter 5), derived using the ratio of the SHAP values for the climatic covariate's influence on predictions and the human mobility covariate's influence.

Translation from evidence to policy

Given these theoretical advances, it is natural to question how best to use the methods and outputs of our retrospective analyses in real-time for informing policy. In each setting of Chapters 4 and 5, there are implications for targeted epidemic control measures. These could be public information campaigns months before a likely large dengue season or targeted vector control measures based on the neighbourhoods that contribute most to onward transmission. As an example, a public information campaign could target reductions in human-vector contact based on high ICEN values that suggest El Niño events and increased outbreak risk.

Looking forward, as observed in the reflections of inequality in the COVID-19 pandemic[699], any model-implied interventions will be more feasible, economical and effective if spatiotemporal models and analyses continue to be developed in collaboration with local data collectors, researchers, and public health authorities. This can be as simple as ensuring that we incorporate all relevant and available data or ensuring that we produce outputs and insights with the timeliness and resolutions (temporal and spatial) required for local decision-making[103,199]. In Chapter 4, the embedding of our models in an early warning system (EWS) was suggested by i) our retrospective analyses with delayed climatic effects from up to four months ago and ii) our short-term probabilistic forecasts with moderate-to-strong predictive performance across spatial scales at forecast horizons up to one month. The forecasts helped validate our models and retrospective analyses, and vice versa, and suggest potential of an operational tool that supports decision-making for different policy/research questions; from retrospective understanding to future-looking forecasts (see below and Figure 4.7).

Indeed, a forecast is only useful if it promotes the decision-maker to take appropriate action, and such control or

surveillance strategies can only be designed using an understanding of *why* a pathogen spreads. Forecasting is just one component of a model's capabilities, and epidemic forecasts have fundamental limits as an unconditional prediction of a future that may be modifiable by policy changes (see Chapters 2 and 8). As we are further challenged by incomplete knowledge of the evolving epidemic dynamics, decision-makers often require timely, quantitative answers to many more questions to make important public health decisions[3,103,158]: i) How many DENV infections are there in the population right now? We could use our learned parametric relationships, spatiotemporal patterns, and methods more generally in downstream analyses for nowcasting models[114,170,540] or to fill temporal, spatial, and population gaps in epidemiological surveillance[69]. ii) How fast is DENV spreading? We could incorporate our insights about epidemic drivers and dynamics in an approach to estimate $R(t)$ (see below). iii) How many cases would we expect next month or next season under specific temperatures and precipitation levels? Understanding past epidemic drivers should enable more reliable and useful scenario predictions and/or seasonal forecasts that public health authorities require in many endemic countries (e.g., Brazil).

As conversion of modelling output to policy-relevant evidence rests on data quality, we note that estimated probabilistic climatic/mobility effects and any downstream uses rest on flawed data. These imperfections stretch across reported cases, human mobility, climate, and demographic data. For dengue case data, flaws include a lack of serotype-specific and age-stratified case data, reporting delays, and representativeness (see Section 2.4). Reporting delays, time interval censoring, and dynamical biases will adversely affect the timeliness and accuracy of our insights and, in a worst-case scenario, mislead policy. For example, if cases were reported in bulk, we could have misestimated delayed climatic effects. Correcting for reporting delays could use recent latent variable methods[113] and/or additional data streams such as hospitalisation, wastewater, or seroprevalence data. Correcting for imperfections of climate reanalysis data and human mobility data would also likely involve additional data such as local weather station data paired with bias correction methods and different types of human movement data such as flight and phone data to capture different aspects of human movement.

Causality, serotypes, and reporting rates

As discussed throughout the thesis, establishing causality in epidemics is non-trivial. We are challenged by partial observations, nonlinearity, stochasticity, spatial structures, and feedback loops. The measured outcome generally does not coincide with the unobserved variable of interest. For example, incident cases of dengue from passive surveillance are not equal to incident DENV infections. Due to data availability, we made simplifying assumptions about causality, as we modelled the direct effects of climate/human mobility on reported *cases* and not DENV *infections*, only indirectly accounting for confounders via random effects (Figure 2.1 E). This means that varying reporting rates, mosquito populations, and other influences across time and/or space could bias our estimates of causal climate/mobility effects on DENV infections. We tried to guard against the effects of such mediators by developing multi-model explanatory and predictive approaches involving: i) cross-validation such as rolling and expanding window model fitting, ii) spatial and temporal random effects in models, iii) model checking with information and predictive criteria, iv) sensitivity analyses such as models with and without climatic/mobility variables, and v) exploring proportions of probable versus confirmed cases across space and time. Adopting these extensive explanatory and predictive approaches across different modelling disciplines should enhance the causal robustness of our evidence synthesis.

Although our approaches can indirectly account for changes in serotype circulation and the age distribution of cases, these could also proxy many other time- and space-varying effects, such as mosquito population counts or mosquito dynamics and genetics. The lack of seroprevalence data or age-and serotype-specific case data meant that we could not *directly* model population immunity, serotype competition, and reporting rates, all of which can control the shape of case-derived epidemic curves[112,244,574]. Thus, this could impact the extent to which our estimates reflect the causal effects of climatic and mobility drivers on the true interacting serotype-specific epidemic curves of infections. Similarly, we could not model the influence of water storage practices that our collaborators in the Peru CDC suspected as a factor in dengue

outbreaks in northern Peru. In short, the effects of climatic conditions and human movement on DENV infections could differ by different immune landscapes, different serotypes, and different human behaviour, and estimated effects could be misrepresented by using the measured outcome (dengue cases) under different reporting rates.

Correcting the models for such variations is non-trivial, likely requiring further data collection. Ideally, we would like large-scale and high-resolution serosurveys over time (see Section 2.5). Although increasing in frequency globally[315], these are unfortunately limited in many dengue-affected regions such as northern Peru or Fortaleza. Serosurvey data could be used to fit models jointly to age-stratified case, climate, and seroprevalence data. This could then allow us to reconstruct FoIs and DENV transmission history, quantify the immunity landscape and calibrate/infer reporting rates across space and time[189,310,314,322]. Focusing on the reporting rate objective, as serosurveys are often challenged by logistical, political, and economic barriers, two key serosurvey design questions then arise: i) What intensity and sample size of longitudinal study do we need to infer reporting rates reliably? ii) What are the relative benefits of these studies over more cost-effective and similarly useful age-stratified case surveillance[244]? To address such questions and quantify hypothetical information gains (e.g., using Value of Information, VoI[700]), a simulation study could be developed and incorporate DENV-specific factors such as circulating serotype(s), age-specific vaccination strategies, assay sensitivity, and individual infection risk factors. By showing what serosurveys *could* offer for inferring true dengue burden and epidemic indicators in simulated settings where we know the truth, and at what level/cost, this simulation study could encourage more and better-designed seroprevalence studies and age-stratified case data collection in resource-limited settings.

Practically, this information is useful for policy. Key lessons from the COVID-19 pandemic were that such scalable and sustainable data collection is vital to address policy questions in public health crises and that cost-benefit analyses should be used to quantify relative benefits of data streams[19]. Having this theory for study design, alongside our various models, in place ahead of time should promote future cost-effective and representative data collection. Furthermore, by using models to develop a more accurate understanding of historical transmission patterns and epidemic drivers, we can work with public health authorities to improve real-time monitoring and EWS systems (Sections 10.2 and 10.3.1, respectively), to quantify the burden of dengue more accurately, and to design safer and more effective vaccination strategies, where appropriate[159,229,244,701].

Mechanisms of mosquito-borne transmission

Another limitation of our retrospective analyses is that we did not directly model the role of the vector in DENV transmission due to the scarcity of relevant mosquito surveillance data (e.g., mosquito densities over time) to parameterise such a mechanistic model. This could limit the epidemiological realism of our models' outputs. Although there exist many alternative approximations such as mosquito suitability indices and mapping estimates, incorporating these risk metrics at the "correct" spatial resolution is non-trivial (see[104,218] and Section 2.4). Previous research has found a poor and nonlinear correlation between these metrics and reported dengue cases[702].

To some extent, our estimation of dynamics at different spatial scales and retention of a limited but causal structure (via climate/mobility-cases relationship) provides a safer approach in the absence of reliable mosquito data to control for an unobserved mediator in the relationship between climate/mobility and DENV infections. We would ideally use time-varying data from enhanced mosquito surveillance networks in dengue-affected settings. Recognising the logistical and economic challenges, particularly in resource-limited settings, we also advocate for community-led initiatives that empower individuals in areas that ultimately suffer the burden of dengue[78,261,703]. For example, ZanzaMapp is a local vector monitoring and decentralised surveillance network in Italy[263].

Contextualising our results further, we identified largely similar climatic risk profiles across spatial scales, although spatial resolution remains a limitation for our analyses in Peru. This could be problematic due to heterogeneous mosquito dynamics and genetics, environmental influences, and human behaviour within the provinces studied. For example, *Aedes*

mosquito have limited ranges and mosquito-borne transmission is often highly local due to location of egg sites and host feeding sites, which can result in heterogeneity in transmission across spatial scales[218]. Ideally, we could use finer spatial resolution data, as was the case for our study on human mobility in Fortaleza. This mobility study consequently provides more targeted and possibly more accurate information for interventions such as vector control, thus enhancing the likely cost-effectiveness and efficacy of interventions.

Operationalisation and generalisability in other resource-limited settings

In Chapter 4, we developed our workflow for climate-informed modelling to be an open-source computational tool for reproducing similar analyses in other geographies. As dengue incidence increases globally and DENV spreads to new areas and more climate-sensitive pathogens emerge, such a computational tool should be increasingly useful. For example, autochthonous dengue cases are increasingly observed in Mediterranean Europe[71,72,355]. Our methods are likely generalisable in these settings with minor changes (e.g., for species and population differences). In this thesis, we focused on developing insights for policy and capacity-strengthening in specific resource-limited LMICs of Latin America (Peru and Brazil). This is because the dynamics and drivers of epidemics are less well understood and surveillance can be challenging in such regions. Moreover, our studied countries have also experienced some of the most substantially increasing dengue burdens globally in recent years[42].

Meanwhile, in Africa, all four DENV serotypes are known to circulate. Case reporting is often sporadic and other surveillance methods are severely limited due to lack of resources (e.g., testing capacity)[261,315]. However, increasing evidence points to an underestimated dengue burden, rising population levels of *Aedes* mosquitoes[704], and silent unreported dengue epidemics[71]. Uncovering the extent and causes of this burden is vital for positive public health outcomes worldwide. Our methods, conditional on funding for case surveillance, genomic and/or seroprevalence data, could be applied to i) uncover environmental and human drivers, ii) fill gaps in spatial and temporal surveillance coverage, iii) develop joint modelling/inference approaches across data streams (e.g., phylogenetic or serocatalytic models alongside models of case incidence[301,313]) and iv) answer predictive questions of DENV invasion (see below). We again suspect that climatic drivers on DENV infections will be both time- and space-varying due to varying human and biological factors, and the calibration of reporting rates will be crucial to understand epidemic dynamics across many African countries.

Joint inference of epidemic drivers

Across Chapters 4–5, we synthesised evidence of environmental and human effects on dengue epidemics, yet research from the COVID-19 pandemic underlined the benefits of joint inference for epidemic drivers/dynamics[15], often using multiple data streams (e.g., COVID-19 NPI, case, and death data[705]). Our Bayesian models included surveillance, climate or mobility, and demographic data. However, in Chapter 4, the climate-based Bayesian hierarchical model only approximates human mobility effects via spatial autocorrelation effects (using a Besag-York-Mollié model[519]), yet in our wavelet analyses, we identified amplifying effects of pairwise populations and shorter distances on epidemic similarity (using a gravity approximation to mobility).

Conditional on obtaining representative human mobility data for Peru or using new and improved mobility approximations[595], we could explicitly include human movement in climate-informed models using the mobility-weighted sum formulation used in Chapter 5. We are not the first to consider the idea of joint inference for orthoflavivirus outbreaks, as the joint effects of climate and mobility have been estimated as determinants of i) DENV spatial invasion and epidemic timing in Pakistan[85], ii) DENV expansion in Vietnam[79], and iii) yellow fever spatial invasion in Angola[265]. We are currently working with collaborators in Peru, Panama, and Chile to extend these climate-mobility modelling ideas to Latin America, and work more generally on uncovering time- and space-varying influences beyond individual roles for invasion, expansion, and timing of dengue outbreaks.

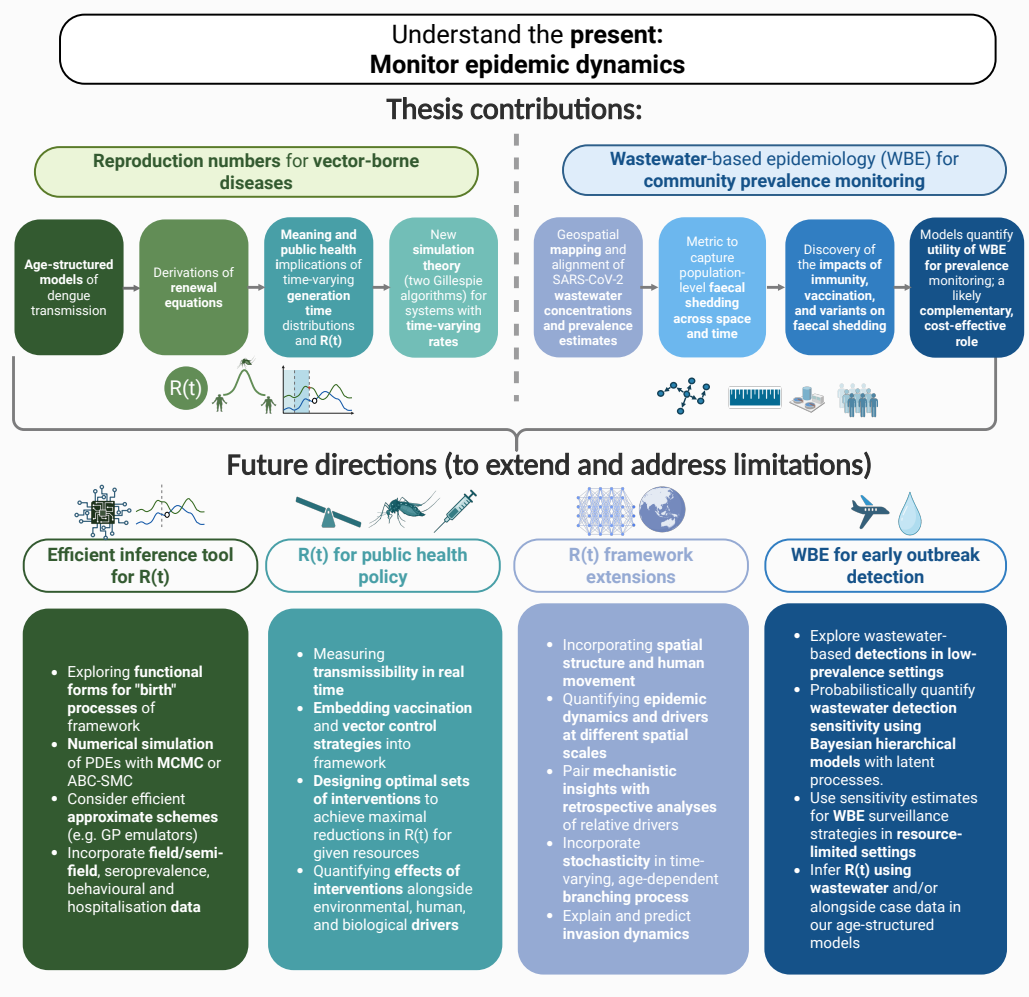


Figure 10.3: **Understand the present:** In Chapters 3, 6, and 7, we introduced methods and analyses to monitor dynamics for COVID-19 and dengue epidemics respectively. Contributions ranged from methods for wastewater-based monitoring of SARS-Cov-2 to theory for estimating reproduction numbers for vector-borne diseases. We summarise here the theoretical and practical contributions (in further detail than the middle column of complementary Figure 10.1), alongside the future directions to extend our methods and address limitations.

10.2 Understanding the present: Monitoring epidemic dynamics

For effective public health policy, timeliness, and reliability are essential[17,165]. However, surveillance data are delayed and incomplete measures of *why*, *when*, and *where* a pathogen is currently spreading (see Section 2.3). To inform decision-making, work towards epidemic control and achieve better public health outcomes, authorities need timely and reliable estimates of epidemic dynamics[17]. These estimates include the early outbreak dynamics, changes in real-time transmissibility, and the impacts of interventions or different epidemic drivers[15].

10.2.1 Reproduction numbers for mosquito-borne diseases

Theory for renewal equations, R(t), and simulation

Measuring how fast a pathogen is spreading (i.e., transmissibility) is crucial. Although reproduction numbers received possibly outsized public attention during the COVID-19 pandemic[107,136], the instantaneous reproduction number $R(t)$ remains an interpretable (at the societal and decision-maker level) and potentially useful quantity that i) enables comparison

of epidemic dynamics across space, time and data streams, ii) provides a time-varying measure of epidemic controllability, and iii) ultimately can directly inform real-time outbreak responses for achieving epidemic control [134].

However, for vector-borne pathogens, theoretical difficulties in estimating $R(t)$ occur due to multiple stages of the human-to-human transmission cycle. For DENV and other mosquito-borne pathogens, such stages are sensitive to environmental, biological, and human influences that vary over time (Chapter 2). This means that there is no fixed distribution of times between primary and secondary human infections – the generation time distribution is time-varying. The renewal equation, the basis for most contemporary real-time methods for estimating $R(t)$ in directly transmitted diseases, is therefore not directly applicable. These theoretical difficulties have practical consequences for our ability to inform policy, as real-time applications of estimating $R(t)$ for mosquito-borne pathogens are rare (see below and [83,377]).

In Chapter 6, we derived renewal equations from a new age-structured system of pathogen transmission across infected and infectious human and mosquito states, where ages represented the times since humans or mosquitoes have entered a particular state. Our renewal equations describe how the incidence of new human infections is generated from previous human infections, based on $R(t)$ and a generation time distribution that varies over time. Our framework is based on the seminal work of Lotka [706] and Kermack and McKendrick [330] and extends the work of Fraser [137] and Cori et al. [139] and subsequent frameworks [140,148,149] to the setting of vector-borne diseases.

Two key advantages of our theory are: i) renewal equations that allow for arbitrary forms of vector-borne pathogen transmission, and ii) the generality of time-varying processes that can account for multiple influences on the stages of the human-to-human transmission cycle. For DENV, this theory can capture influences such as the behaviour of *Aedes* mosquitoes and humans, thermal dependencies across the mosquito and virus, and further interacting environmental effects on mosquitoes. As demonstrated in Chapter 6, many time-varying variants and extensions of the ubiquitous Ross-Macdonald assumptions [104,121] could be incorporated, ranging from mosquito population carrying capacities to human population counts. Our approach differs from the few previous attempts to derive $R(t)$ for DENV [156,375,376,377]. These attempts relied on generation time distributions that are fixed over time or respond only to changes in current temperature without accounting for the many historical and current conditions that affect stages of the transmission cycle. This difference can have practical consequences as misspecified generation time distributions can yield incorrect insights for policy decisions about epidemic control (e.g., Is $R(t)$ above/below threshold?) [107,134,142]. $R(t)$ and the generation time distribution are intrinsically linked, and without our mechanistic formulae, these quantities would be unidentifiable. Our methods that can correct potentially inaccurate assumptions in formulae for generation time distributions and $R(t)$, which could be of great public health importance.

A more general but related benefit is that we provide the meaning of, for a vector-borne disease, a generation time distribution that varies over time (t). It represents the relative/proportional contribution of different combinations of sub-population (here, infected and infectious humans and mosquitoes) ages to the expected number of new human infections generated by humans infected at time t throughout their infected lifetime should conditions affecting human transmission remain the same. Note that due to the multiple stages required for a human-to-human infection, there is a dependence on historical conditions before time t – the sub-population ages in the generation time distribution are the values at the times at which each sub-population played its role in the transmission cycle (see Figure 6.1 C). This complex, lengthy and careful mathematical construction of a generation time distribution that varies over time is nuanced and possibly paradoxical – we define both the generation time distribution and $R(t)$ at time t to depend on *historical* conditions yet with respect to conditions affecting human transmission at time t remaining the *same* after time t. This apparent conflict remains mathematically proper, aligns with the theory for directly transmitted diseases [145], and simply reflects that both the generation time distribution and $R(t)$ are *counterfactual*, *instantaneous* measures of relative contributions to human infections and of time-varying transmissibility, respectively.

With suitable adaptations, our concepts for the time-varying generation time distribution and $R(t)$ are applicable to

other infectious diseases that require a modeller to track times between infections using stages of the transmission cycle. While it is advantageous that our framework requires no assumptions about the drivers of time-varying processes (e.g., avoiding exclusive temperature forcing), the drawback of the framework's generality is that functional or prior distributional forms for the time-varying processes are not obvious.

To further understand our system, the importance of stochasticity, and how to parameterise the processes that control infection dynamics, Chapter 7, we developed new simulation theory. Our theory focuses on exactly simulating stochastic reaction systems that experience time dependencies in the propensities at which events occur. Our theoretical advances consist of two exact simulation algorithms that extend the popular Gillespie direct and first reaction algorithms[50,51] that are commonly used to simulate systems in chemistry, biology, finance, and many other branches of science. These original Gillespie algorithms implicitly assume that, between two events, the propensity at which events occur is fixed, as no events can have occurred to change the system. We extended these direct and first reaction algorithms to allow for external time-varying influences that act on event propensities between events. Using simple chemical reaction systems, we validated the theory by comparing population trajectories from the two algorithms with each other and with numerical ODE solutions. In a more elaborate example, we used this theory to simulate an individual-based model of DENV transmission, based on the equations proposed in[48]. Our algorithms benefit from simplicity in implementation and are most suitable for simulating small stochastic systems. Like any Gillespie method[615,618], these algorithms are likely prohibitively expensive for large systems. On the other hand, stochastic methods are generally more important for us in smaller systems such as at early outbreak stages or in small population count settings. Our algorithms improve upon approximate routines (e.g., using central limit theorems[625]) by enabling exact simulation and avoiding approximation error. Our simulation theory is more general than our specific applications. Our methods can be used if faced with time-dependent propensities between events when simulating stochastic models of dengue or other infectious diseases and, more generally, small stochastic systems in any branch of science.

R(t) for targeted public health policy: Controllability, interventions, and epidemic drivers

Given this mathematical and statistical theory, it is natural to again question how this theory addresses the public health priorities of the thesis (Chapter 1). Due to the lack of theory for understanding how to generalise methods developed for directly transmitted infectious diseases, modellers have rarely been able to provide reliable and mechanistically driven estimates of $R(t)$ to inform public health policy for dengue in real time[83,377]. In Chapter 6, we provide theoretical advances which can support three key public health activities: i) measuring time-varying epidemic controllability, ii) understanding mechanistically the impacts of different interventions and epidemic drivers on transmissibility, and iii) enacting more targeted and effective DENV public health policy.

First, we derived formulae for epidemic controllability in the early stages of the outbreak (basic reproduction number for humans, R_0) and throughout the outbreak ($R(t)$). Estimates of R_0 should inform long-term resource planning strategies by informing whether a mosquito-borne disease is likely to cause a major outbreak and, if so, the likely outbreak size. Focusing on $R(t)$, this is also measured at the *individual* level of a human, as opposed to reproduction numbers for systems of transmission. It provides decision-makers with almost identical information/meaning as the familiar $R(t)$ for directly transmitted infectious diseases – the expected number of new infections generated by an infected individual throughout their infected lifetime, should the conditions affecting human transmission remain the same as at time t. Conditional on the statistical literacy (see below) and uncertainty quantification that are vital to translate scientific evidence into policy[18,105,141,203], we can still provide decision-makers with a familiar, interpretable quantity that has the useful threshold property: $R(t) > 1$ still indicates a growing epidemic of human infections. Furthermore, $R(t)$ can be compared across different regions, different stages of the epidemic, and different data streams[134]. Consequently, our theoretical advances should enable real-time measurement of changes in pathogen transmissibility, which leads to the second key benefit –

understanding the impacts of different interventions and epidemic drivers.

We have not considered optimal vector control strategies in this thesis – see [707,708,709]. However, in our mathematical framework of Chapter 6, we could embed different public health interventions that target reductions in transmissibility at different stages of the DENV transmission cycle. Throughout our systematic review of the COVID-19 pandemic [54], we found that it was often through $R(t)$ that modellers estimated the real-time effectiveness of interventions. Despite limited evidence of effectiveness in controlling outbreaks, costly interventions for dengue (e.g., vector control by insecticide spraying) are still being implemented globally [71,90,91]. Although an effective DENV-specific vaccine has long faced many challenges, there are now three leading vaccines at varying stages of implementation. Dengvaxia and Qdenga have been licensed for use in limited settings, while the TV003 tetravalent vaccine is at phase three stage of clinical trials [71]. These vaccines remain challenged by logistical and economic barriers and vary in efficacy. Other novel strategies such as Wolbachia-based vector control have been demonstrated to reduce dengue cases and hospitalisations, but their effectiveness will always be difficult to quantify in observational outbreak settings as these population-level indirect control measures are inherently confounded [71,90,94,710].

Looking forward, designing optimal sets of combined vaccine and vector control strategies will be essential for long-term and cost-effective DENV control [569,711]. These strategies will involve traditional, novel, and/or as yet unknown methods and will be complemented with short-term interventions such as public information campaigns to reduce human-vector contact. In our framework, we can again embed interventions previously tested in experimental and/or observational settings. We can then design optimal integrated strategies (see below) to achieve maximal reductions in $R(t)$, subject to the decision-maker's resources for the interventions. Focusing on intervention effects in our framework, current or future DENV vaccines could decrease probabilities of human infection upon infectious mosquito bite and/or of mosquito infection upon bite of an infectious human. Wolbachia strategies could reduce the density of the (female) mosquito population by increasing death rates. Public information campaigns could reduce human-vector contact via lower biting rates.

However, effective implementation of strategies will rely on *quantifying* (likely via inference) these hypothesised effects using their impacts on epidemic dynamics in real time outside idealised, controlled environments [71,90]. These interventions will likely have differential effectiveness in different populations in different epidemic phases where multiple interventions may be in place, many confounders exist, and dynamics are highly nonlinear and stochastic (see Chapter 2, [429]). Like any statistical analysis, effects cannot be *directly* measured. However, we can estimate the effectiveness of interventions for reducing $R(t)$ by fitting our model to empirical data and using prior distributions for different effects of interventions in our inference framework (see below). As illustrated by the usage of $R(t)$ throughout the COVID-19 pandemic (e.g., [269,705]), this procedure would enable both greater scientific and policy understanding through empirical estimates of policy effectiveness. Estimating these observed effects in the population is not a substitute for within-host models, ABMs, gold-standard cluster randomised controlled trials (e.g., for Wolbachia [94]), or necessary vaccine clinical trials (e.g., [193]). However, during epidemics, experiments and trials are challenged by ethical barriers and the timeliness required to control epidemics that are constantly evolving and often poorly understood. Here, we are instead considering post-implementation surveillance [707]. The COVID-19 pandemic and other outbreaks (e.g., Ebola in Liberia) [15,712] have highlighted that to inform decisions and maintain public and decision-makers' trust, it is necessary to quantify, in a timely and reliable manner, the impacts of interventions targeted at both the population and individual level.

As an example, suppose that the reported incident cases and estimates of $R(t)$ have decreased, with delay, after the implementation of control measures. There will likely be both scientific and public disagreement about *why* the epidemic is now under better control. Having our mechanistic structure and well-defined formulae defined in Chapter 6 will allow researchers to perform joint inference and probabilistically quantify the most plausible theory/reasons based on current available evidence. This mechanism-driven inference approach i) guards against misleading insights caused by non-identifiability as inference algorithms are guided by mechanistic formulae, and ii) extends beyond simple and potentially

incorrect regression or correlation analyses that infiltrated research during the COVID-19 pandemic (e.g., using $R(t)$ estimates in downstream analyses without considering uncertainty and upstream initial assumptions). Our mechanistic framework is advantageous in the sense that $R(t)$ would *not* be directly inferred, but instead defined in a mechanistic way. Conditional on clear communication of intervention effectiveness, this mechanism-led inference approach should help to build better public trust. Since William Mompesson's strategy of voluntary quarantine during the Eyam plague in 1665[713], public discontent has naturally persisted about control measures, and balancing ethical considerations[683] against quantifiable effectiveness will be important for maintaining and enhancing public trust in both models and policy.

Meanwhile, we can also estimate the impacts of the many human, environmental, and biological drivers/influences on dengue epidemic dynamics at different times and in different locations. We could embed these influences using current and historical covariates for the time-varying processes of our framework. This allows us to extend beyond assumptions of fixed parameter relationships/mechanisms that often only use laboratory-derived thermal performance curves. Instead, we can handle time variations more generally and could *jointly* quantify the impacts of multiple interacting epidemic drivers on $R(t)$. For example, we could quantify the interacting effects of temperature and precipitation alongside human and biological confounders such as population immunity or serotype circulation. Here, we can measure performance using information criteria and scoring rules for different models with different functional forms involving different combinations and parameterisations. Each of these models can reflect different roles of epidemic drivers, and their relative importance could be inferred by the measured appropriateness of each model. Given the likely many plausible functional forms for hypothesised drivers, we suspect that AI approaches such as foundational time series models could be useful to constrain the function space. Testing different assumptions is especially important in the unfolding climate crisis, as there are many unknowns in how the commonly used thermal performance curves extrapolate to the wild[83,154,155], and these curves are the basis for projections of the future incidence and distribution of mosquito-borne diseases under different climate change scenarios (see Chapter 6).

Ultimately, the third advantage is a coalescence of the first two: providing information for more targeted and effective public health policy. The decomposition in Chapter 6 of $R(t)$ into reproduction numbers between stages of human-to-human transmission enables more detailed analyses of *why* transmission is occurring – which components of the transmission cycle are driving a growing or shrinking human outbreak. For example, suppose that we have a currently growing epidemic. We can inform strategy design by estimating, for the current $R(t)$ and the state of the epidemic, which of the different contributors to transmissibility should be targeted to achieve 'optimal' reductions in $R(t)$. Optimality conditions will need to be pre-defined and could be the maximal reductions subject to the decision-maker's given budgets. Feasible control strategies may be demographic-specific human vaccination, Wolbachia-mediated or Wolbachia-infected mosquitoes, and indoor/outdoor spraying strategies. By estimating the historical effects of these strategies on different contributors to $R(t)$, the optimal sets of cost-effective interventions could be designed, including the intensity/scale required to achieve different levels of reductions in $R(t)$. As estimated optimal decisions will be transient, subject to uncertainty, and specific to a decision-maker's preferences, embedding the preferences and rewards in a reinforcement learning or model predictive control framework could enable iterative updating of implied optimal decisions for interventions in real time. Adopting a forward-looking approach, we could then further inform policy by developing scenario predictions under different assumptions/understanding of intervention effectiveness and community responses.

Computational inference tool and data for estimating R(t)

In an exciting time for dengue control strategies[71,90], we view our theory in Chapter 6 for real-time epidemic control as a necessary first step towards a useful computational inference tool for estimating $R(t)$. The most critical limitation is the lack of real-time applications, and providing many of the benefits above is constrained by i) our limited knowledge of how best to perform inference with this framework and ii) insufficient data.

We have partially addressed the former constraint with our new simulation theory in Chapter 7, which can help to parameterise prior distributions for time-varying processes and understand the likely impacts of stochasticity. A computational inference tool can only provide timely and reliable information if it is both efficient and accurate. The ability to recover, together with uncertainty, the simulated data will be a vital first test of any inference tool. This will determine the certainty of evidence statements communicated to decision-makers about epidemic control.

In future work, we could fit our model to reported incident case data using separate observation and dynamic processes for cases and infections, respectively. An inference method could pair this observation process with numerical simulation of PDEs using Bayesian tools such as MCMC or ABC-SMC[714,715,716]. If this scheme proves unstable or more likely inefficient, we could consider GP emulators (a class of AI surrogate models), which rely on a smaller number of simulations of the PDE and permit probabilistic estimates of parameters. Alternatively, we could use AI and hybrid approaches such as PINNs[717] and symbolic regression[718]. PINNs may struggle during training due to the likely limited and noisy data, while generalisability benefits of symbolic regression could be outweighed by scaling difficulties when we extend to higher dimensions (e.g., with spatial structure).

With respect to data for inference procedures, it was mosquito field trial data that allowed researchers to first test the well-known Ross-Macdonald theory for mosquito-borne diseases[719]. Similarly, for us today, such field and semi-field data, along with laboratory and mosquito surveillance data, could help to i) calibrate components of the time-varying processes of infection and ii) estimate the underlying female *Aedes* population. For example, we would like to explore how commonly-used mosquito and viral traits from the laboratory respond in the wild[83]. Recent evidence calls into question their direct extrapolation from the laboratory to the wild and points to local *Ae. Aegypti* adaptation and/or thermoregulation[154,155]. Furthermore, fluctuations in diurnal temperature appear to influence *Ae. Aegypti* DENV susceptibility and death rates[215]. Here, we are aided by the move towards standardised, global databases of laboratory-based experimental results (e.g., temperature-dependent *Aedes* mosquito traits[720]).

While we, as researchers, often focus on the need for more high-quality mosquito data when modelling mosquito-borne diseases, humans remain critical for transmission. The idea of embedding human *behaviour* into mechanistic infectious disease models is not new[199,721,722], yet the indirect transmission for mosquito-borne diseases adds further complexity. Ideally, data on human-mosquito contact patterns such as host feeding preferences[723] and context-specific human behaviour[199,721] would inform more realistic models of transmission. This would be particularly important at more granular spatial scales. Compared to incorporating human mobility patterns, modelling individual-level characteristics is more challenging, yet ABMs paired with large language models may offer routes to modelling complex human behaviours[289].

Once again, seroprevalence, genomic surveillance, age-stratified case, and hospitalisation data could all improve the implementation of a possibly serotype-specific inference tool. Seroprevalence and/or hospitalisation data will be especially important for estimating reporting rates and transmission patterns in areas with historically poor surveillance (e.g., Africa) or invading dynamics (e.g., Europe). Hospitalisation and healthcare availability data[724] could improve the estimates of time- and space-varying case reporting delays. Integrating these heterogeneous data streams is probably mathematically simple, compared to collecting data, which is often logistically and economically challenging for dengue-afflicted regions[176,703,725]. Given the promise for our methods of providing outputs that can inform epidemic control strategies, we hope that our theory, computational infrastructure, and collaborations can promote this data collection.

Spatial considerations for R(t)

Our framework, like many renewal equations[15], is currently limited by the lack of an explicit spatial structure and therefore assumes homogeneous mixing for both humans and mosquitoes within a region. Practical implementations will need to consider and investigate the effects of spatial scale, due to the aforementioned heterogeneities in transmission across spatial scales[190,195,218]. Similarly, considering multiple regions simultaneously is likely important, as human movement is clearly

a determinant of dengue epidemic patterns, while estimated climatic effects can be confounded by such movement[232].

One way to incorporate spatial structure in renewal equations is to include time-varying infection importations as a background term, thus leading to a Volterra equation of the second kind for the incidence of new human infections $i(t)$ (see[129,371,726] for directly transmitted diseases):

$$i(t) = m(t) + R(t) \int_0^t i(t-\tau) w(t,\tau) \mathrm{d}\tau \tag{10.2.1}$$

where $m(t)$ is the background term that describes imported infections that contribute to onwards transmission in the studied region, $w(t,\tau)$ is the generation time distribution that varies over time with τ being the time between two human infections and $R(t)$ is the instantaneous reproduction number. Having separate background and internal processes induces additional complexity and identifiability issues for inference. Similar to our derivations of Chapter 6, we would need to derive eq. (10.2.1) starting from a spatially explicit age-structured framework which extends our original models. Practically, we would then have two competing processes, both of which could have generated the incidence of new infections at time t.

We could infer $m(t)$ and/or its components for each region as a latent process – possibly parameterised as a function of all regions' current and historical incidence levels, spatial connectivity covariates (e.g., mobility approximation and/or data), and/or other region-specific covariates/traits (e.g., mosquito populations). Our mechanistic structure and associated formulae such as generation time distributions, for the two processes, again provides clearer information and reduces identifiability issues for an inference routine to appropriately weight and distinguish the fractions of infections that stem from each of the two competing processes.

There is no necessary dependence on human mobility data in eq. (10.2.1), yet the availability of such data could help to address any residual identifiability issues and incorporate realistic patterns of human movement. A recent extension of the renewal equation for directly transmitted diseases includes spatially explicit components and connectivity between regions determined by human mobility data[727]. This could be a different starting point for future extensions. In any of these extensions, we can use our generation time distribution to estimate the timing of delayed impacts of human movement, as a commuting/migrating infected human can only generate new human infections after the multiple stages of human-to-human transmission.

Conditional on incorporating the spatial structure into our framework, we can further quantify different epidemic drivers across different regions and stages of an epidemic. For example, we could compare whether the estimated relative importance of human movement or climatic conditions for driving pathogen spread (e.g., $R(t)$) aligns with our retrospective analyses that use multi-model approaches (Chapters 5 and Section 10.1.1). Similarly, public health authorities in LMICs often have fixed and very limited resources for interventions. By identifying which regions could be targeted to achieve maximal reductions in current transmissibility, we could enhance the cost-effectiveness of resource allocation. We can again use a triangulation-like approach by comparing these allocation insights i) with the policy implications of source-sink dynamics of Chapter 5 and ii) with the implied optimal decisions of our new evaluation framework (Chapter 8) paired with our forecasting models. As we argue throughout the thesis, this development of scientific understanding and public health insights from multiple models and analyses should strengthen their robustness and utility.

Stochasticity, climate change, and invasion dynamics

In this thesis, we have not explicitly modelled invasion dynamics or spillover events, but DENV is expanding globally in distribution, within and across nations, and other arboviruses are emerging or re-emerging in new areas (e.g., Oropouche virus)[38,63,355,356,728]. Such invasions globally require modelling approaches that handle the likely low counts, either of infected/infectious humans and mosquitoes or of humans and mosquitoes more generally. Despite these expanding public health threats, the currently deterministic framework of Chapter 6 is less suitable in such areas with likely stochasticity.

Estimating transmissibility is essential to understand the likelihood of a future major outbreak[15]. Meanwhile, understanding why past invasions and large outbreaks have occurred is essential for more accurate predictions of *when* and *where* DENV will invade in the future[227]. There are various ways to incorporate stochasticity in our framework, including stochastic McKendrick-von Foerster equations that use general additive stochastic noise in transition rates. This framework is limited by the currently known approaches to solving these stochastic PDEs. These approaches use a likely inappropriate compartmentalisation (for infectious disease dynamics) of populations into juvenile and adult sub-populations, which simplifies the equation to a stochastic functional differential equation[729]. We could alternatively assume a specific noise term such as a Brownian motion or a fractional Brownian motion. This approach would require further investigation as it would likely require re-interpretation as an abstract linear equation in a Banach space[730,731].

The most likely natural approach to include stochasticity is to derive renewal equations that originate from more general time-varying branching processes such as the Crump-Mode-Jagers (CMJ) branching process[732,733,734]. Pakkanen et al.[152] derived a stochastic branching process for the incidence of new infections using a time-varying CMJ process that allowed the offspring distribution of infected individuals to vary over time. Curran-Sebastian et al.[372] extended the framework to incorporate imported infections. However, this theory was developed for directly transmitted diseases. It is unclear i) how this *individual*-based branching process framework relates to our age-structured framework of multiple *populations* across humans and mosquitoes and/or ii) how or if we can derive corresponding renewal equations that govern the behaviour of any branching process.

The advantages of a branching process model include the ability to estimate the probability of a major outbreak using the probability generating function and the offspring distribution (the number of infections generated per infected host)[735]. Branching process models can also quantify individual-level variation in transmission, the likely size of the outbreak, and the probability of outbreak extinction[15,349], but only age-independent branching processes have been explored for mosquito-borne diseases (e.g.,[736,737,738]). These multitype branching processes implicitly assume time-invariant generation time distributions, which can be unrealistic for many mosquito-borne diseases (see Chapter 6).

Looking forward, working with the authors of[152,372], we are working to extend the general CMJ process to a multitype branching process across human and mosquito populations. Although we do not know the functional forms, our extension will likely involve adaptation of the individual-level reproduction process of Curran-Sebastian et al.[372]. We need to further capture dependencies in the time interval between infections (τ) on historical and current conditions within the time window of length τ. This approach would thereby incorporate i) the necessary historical and current dependencies into the expected number of new secondary infections per infected human and ii) the associated stages and times of this human-to-human transmission cycle.

To understand the public health utility of this branching process approach, suppose that there is a DENV importation on the outskirts of a historically at risk area (as observed for expansion pathways in Brazil[355,356]). Low infection numbers are likely, particularly in the early stages of the outbreak. What conditions are necessary for a major outbreak to occur? What is the probability of such a major outbreak and can we incorporate our previous findings for human movement and climate? What is the probability of further invasion to other neighbouring or connected regions? Branching process models can help answer these questions, and by preparing for potentially small outbreaks, we can enhance control for larger outbreaks[739].

Given our interdisciplinary focus, it would be useful to i) compare the results of branching process models with those of diffusion models[356,740] and other predictive spatiotemporal models[227], and ii) pair the branching process models with genomic and/or wastewater surveillance[741,742] to use detected importations to predict future outbreak dynamics and spatial invasion. The latter additional data streams could be essential, as traditional data streams will likely be challenged in these epidemiologically naive populations.

10.2.2 Wastewater-based epidemiology (WBE)

WBE for prevalence monitoring

In recent years, WBE has been proposed for many surveillance purposes before and during outbreaks[297,491]. However, WBE is rarely a dominant surveillance method despite offering i) cost-effective monitoring relative to other methods (e.g., collation of disease cases) and ii) unbiased testing of population-level epidemic dynamics. Historically, WBE research has concentrated on successfully supporting the fight for poliovirus eradication, and more recently, on the surveillance of COVID-19. The systematic review of Kilaru et al.[297] identifies that most of the research did not link wastewater concentrations to other measures of pathogen transmission. For COVID-19, most analyses were correlation- or model-based investigations of the relationship between wastewater viral concentrations and clinical surveillance data (e.g., reported cases), often for short time periods in a single small community (e.g., university) or wastewater catchment area[297,303,487,743]. Such studies spanned across different wastewater and clinical surveillance programmes and suggested the potential for WBE to be used for early outbreak detection, variant tracking, and small-scale epidemic monitoring. Although WBE could indicate the concentration for different variants of SARS-CoV-2 circulating in the population, it was unclear how WBE could best be used to monitor community prevalence of SARS-CoV-2 in high spatial resolution across a country and over the course of an epidemic.

In Chapter 3, we used two rare datasets enabled by the UK COVID-19 pandemic response – prevalence survey (REACT-1) data and wastewater concentration data[302,308,309]. We performed the first wastewater-prevalence analysis in local authorities of a nation (England) across the COVID-19 pandemic[44]. The methods contained in the analysis included a geospatial population weighting framework that mapped wastewater concentrations from the treatment plant to the local authority level, which may be useful for future outbreaks.

We also introduced a new metric with several implications for public health. This metric monitors how faecal shedding of a virus (here, SARS-CoV-2) in a population evolves over time and varies across space. During the 21-month study period, we found that the metric-implied level of faecal shedding varied according to time variations in circulating variants, population immunity levels, and vaccination coverage. Faecal shedding profiles are generally poorly understood and often assumed to be fixed across time and individuals. However, by revealing these factors in the prevalence-to-wastewater (concentration) relationship, we provide information for the uses of WBE in monitoring epidemics. This should inform public health policy in current and future infectious disease outbreaks, as tracking the identified factors (variants, immunity, and vaccination) will inform when and where recalibration is needed for a wastewater-informed model that is used for surveillance (e.g., a model for cases, prevalence, or $R(t)$). This recalibration is crucial to ensure reliable wastewater-model-based estimates of epidemic dynamics. Our findings capture how our theory can inform triggered approaches to data collection, which are vital for sustainable and useful surveillance during epidemics[175].

Our analyses were constrained by the limitations of wastewater data, including unreported uncertainty in concentration estimates, possible dilution effects, unreported wastewater sampling techniques, and/or treatment-plant-specific conditions (e.g., different distances from the treatment plant)[489]. We were also limited by possible uncertainties[308] and coarse temporal resolution of survey rounds in the prevalence survey data. We tried to address these limitations using sensitivity analyses, different modelling approaches, and extensive data splitting and cross-validation procedures (see Chapter 3).

We developed predictive models that used SARS-CoV-2 wastewater concentrations to estimate community prevalence (our response variable). This allowed us to quantify the utility of WBE for monitoring the community prevalence of SARS-CoV-2 over sustained periods of time and throughout the nation. We found that WBE alone can fill short-term temporal ($<$ 1 month) and spatial gaps in prevalence surveillance. However, we found that WBE is likely insufficient to estimate community prevalence over longer periods of time, potentially limited by the time-varying faecal shedding in communities. Alongside further findings that WBE improved prevalence estimation in reduced-size surveys, we concluded

that WBE could complement, but not replace, costly but highly informative prevalence surveys.

The future of WBE for early outbreak detection

Taking these ideas further, in our WBE analyses (Chapter 3), we tentatively suggested that without concurrent prevalence surveys, early outbreak detection could be an ideal surveillance use for WBE. Here, we will discuss this suggestion, describing how our work can be a starting point for an early outbreak analytic tool that should be developed now to help the early response for future outbreaks[744].

Due to logistical, political, and economic challenges, most countries will struggle to have the same high-quality, large-scale wastewater surveillance and prevalence survey programmes that we explored in Chapter 3. It is currently unknown how best to use WBE and our new methods in such resource-limited settings. Take the familiar example of a pathogen or variant invasion in an epidemiologically naive region with limited resources for large-scale WBE and/or prevalence surveys. Key public health questions often posed are: Given a decision-maker's current budget for surveillance resources, how long until, and how many infections have occurred by the first detection in the wastewater?

In St-Onge et al.[742], the authors considered a wastewater surveillance network of sentinel airports that was designed to optimise resources and maximise early warning ability for different pathogens. For a similar question, our methods for SARS-CoV-2 could be generalised to other settings and pathogens. Focusing on treatment plant sampling, we now consider the reverse direction of our mapping in Chapter 3 – wastewater concentration at a treatment plant level as a function of estimated prevalence. Using our spatial mapping approach and the following Bayesian hierarchical model, we can assign the survey-based prevalence from LTLA level to the treatment-plant-level and estimate the probability of detection for each wastewater sample:

$$y_{irs} \sim \text{Bernoulli}(p_{irs}) \tag{10.2.2}$$

$$p_{irs} = f(w_{irs}) \tag{10.2.3}$$

$$w_{irs} = \beta_0 + \beta_{\text{prevalence}} \cdot z_{ir} + \beta_{\text{plant[i]}} \tag{10.2.4}$$

where i is the treatment plant, r is a prevalence survey round, s is the numeric index for each of the multiple wastewater measurements in a prevalence survey round, y_{irs} is the binary detection indicator for the wastewater measurement, w_{irs} is the corresponding wastewater viral concentration, p_{irs} is the latent process describing wastewater sensitivity (probability of wastewater detection), f is a (smooth) function (e.g., inverse logistic function) mapping to the probability space, β_0 is the intercept, z_{ir} is the community prevalence estimate for the survey round, $\beta_{\text{prevalence}}$ is a parameter to control the effect of community prevalence on wastewater concentration, and $\beta_{\text{plant[i]}}$ is the plant-specific spatial random effect parameter.

By post-processing the latent process for the wastewater detection probability(p_{irs}), we could consider different public-health-relevant axes (or aggregation schemes), and model how wastewater detection sensitivity varies with different prevalence levels, different covariates, time, and/or space. For example, the sensitivity of WBE for detections will depend on the flow normalisation techniques (e.g., to account for rainfall dilution[489]), the studied population, and the pathogen itself, alongside any population (cross)-immunity.

A better understanding of the detection probability will inform i) sampling strategies such as how often to sample, ii) the best use cases of WBE for outbreak detection and low-prevalence settings, iii) surveillance and data collection such as where to target wastewater surveillance, and iv) targeted and timely public health measures such as early interventions.

To mimic resource-limited settings, we could consider reduced-size surveillance designs that satisfy a decision-maker's available budget. We could then model the expected number of wastewater detections from a given number of samples/tests for given prevalence levels across space and time. Using the estimates of WBE sensitivity for SARS-CoV-2 (using eqs (10.2.3) – (10.2.4)), the whole procedure can be repeated in simulated environments for DENV-like and other pathogens (e.g., Ebola-like). This would require simulated data, possibly from a branching process model for early outbreak dynamics,

and faecal shedding and serological profiles for the studied pathogen. The results of the WBE analyses in the simulated environments could help to i) enhance our currently limited understanding of WBE for sentinel or syndromic DENV surveillance, and ii) achieve the hypothesised, yet currently limited, public health value of WBE[261,745,746] for detecting outbreaks of different pathogens in resource-limited settings before epidemic control becomes infeasible.

Other directions for WBE

Our work unlocks future studies for early outbreak detection, yet outbreak detection is just one priority of public health authorities and one use case of WBE, while WBE itself is just one real-time surveillance method (see Chapter 2 and [19,261,297]).

Several modelling approaches have used WBE to estimate $R(t)$ [373,747,748,749], including joint inference alongside reported cases[750,751]. For DENV and other arboviruses, conditional on advances beyond the current abilities of WBE[261,299,580], we could i) compare and identify any time lags between estimates of reproduction numbers and growth rates from our framework in Chapter 6 (which would use reported cases in the observation process) to those derived from wastewater alone, and/or ii) perform joint inference for our framework of Chapter 6 by embedding wastewater concentrations and time-varying shedding profiles (e.g., pairing with reported cases to inform the latent process of total counts of human infections). The setting-specific public health value of incorporating WBE as an additional data stream could again be quantified in a VoI framework by measuring information gains when estimating $R(t)$ or when using estimates of $R(t)$ in a forecasting model for future cases.

10.3 Planning for the future: Forecasting epidemic trajectories

As public health authorities work to control and mitigate the spread and impacts of emerging, epidemic, and endemic pathogens, there is a growing interest in forecasting epidemic trajectories[20,163,170,173,288]. This growing interest has led to advances in forecast modelling, evaluation, and communication. Epidemic forecasting is important as it can inform real-time outbreak response, longer-term public health planning, and data collection[163,170,435]. However, for dengue, there has been limited success in developing generalisable platforms for real-time forecasting of epidemics, while forecast evaluation for infectious diseases more generally has lacked focus on the decision-maker and the ability of forecasts to inform their difficult, uncertain, and consequential decisions.

In this section, we will discuss i) how we can "better" forecast dengue epidemic trajectories and ii) how "better" can be measured as the value/utility of a forecast for different decision-maker priorities and preferences.

Our development of a multi-model framework (Chapter 4) is a stepping stone toward operational climate-informed dengue forecasting in Peru, while our evaluation framework for forecasting of any infectious disease (Chapter 8) shifts the perspective of forecast evaluation from forecasters to decision-makers and measures the value of a forecast by its ability to inform specific public health decisions (Figure 10.4).

10.3.1 Forecasting models

Developing a reproducible, sustainable, and generalisable pipeline

Faced with an unfolding climate crisis, climate-informed forecasting and scenario predictions of epidemic trajectories have never been more crucial and will remain pivotal due to projected increases in incidence of climate-sensitive infectious diseases[37,38,39,40].

The forecasting of any epidemic is inherently uncertain and this art/science remains poorly understood[103,170]. As dengue is generally seasonal and endemic with multiannual cycles[438,539,725], forecasting dengue epidemic dynamics should be a feasible and arguably predictable task, at least for short-term horizons. However, dengue forecasting is challenged by

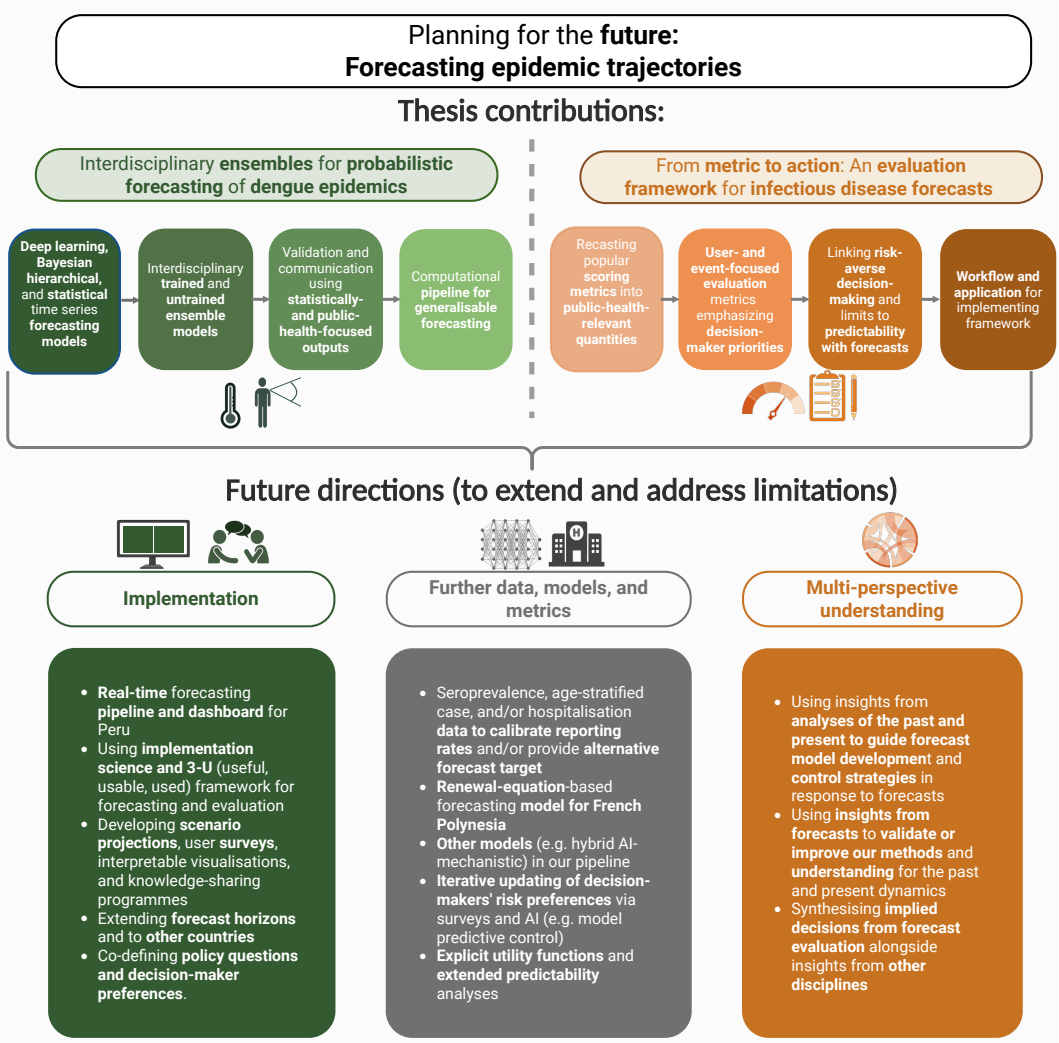


Figure 10.4: **Planning for the future:** In Chapters 4 and 8, we introduced methods and analyses to both predict and evaluate predictions for epidemiological quantities respectively. Contributions ranged from methods for interdisciplinary, ensemble-based forecasting of dengue epidemics to methods for measuring and improving forecast value for decision-makers. We summarise these theoretical and practical contributions (in further detail than the right column of complementary Figure 10.1), alongside the future directions to extend our methods and address limitations.

the many complex human, environmental, and biological factors that are often poorly captured by data. These time- and space-varying influences pose many challenges in developing a forecasting model. Any model must be flexible and able to adapt to such factors, while still having stable performance across space and time to provide timely, reliable, and useful outputs for decision-makers. The direct or indirect incorporation of these factors into a model, together with uncertainty, is exacerbated by the often limited real-time availability and representativeness of data in resource-limited settings where the burden of dengue is greatest [159,278,725]. The most well-known dengue forecast competition raised an open challenge to improve dengue epidemic forecasting, after finding poor performance for short-term forecasts during the early outbreak stages and during years with higher incidence, and consistently better performance for a multi-model approach [169]. A recent competition for seasonal forecasting of dengue epidemics in Brazil found a similar poor performance of individual forecasting models, particularly during years with higher incidence of cases [592].

In Chapter 4, we introduced theory for a new climate-informed forecasting framework. Paired with the retrospective analysis toolkit (Section 10.1), this forms a key component of our proposed multi-model approach for a comprehensive operational decision support tool to monitor and control dengue outbreaks. Without the retrospective analyses, a forecast's

utility could be limited as there could be no obvious actions for a forecast to inform (e.g., where or how to target control strategies). Similarly, understanding the drivers from retrospective analyses and estimating $R(t)$ is more important if it can inform models and proactive control strategies for the unknown future that forecasts target.

We developed our climate-informed forecasting models by recognising the benefits of i) multi-model, ensemble approaches that combine different strengths of individual models, ii) modern deep learning methods such as pre-trained foundational time series models whose training stretches across disciplines and datasets, and iii) traditional yet flexible statistical time series methods that exploit province-specific non-stationarity and temporal signatures. Consistent with many epidemiological forecasting studies [162,170,172], our ensemble models generally outperformed individual models across space and time, when using both statistical metrics (proper scoring rules) and public-health-focused outputs (indicators for outbreak and surge detection). Although we developed this framework for dengue epidemics in Peru, our methods can be generalised to forecasting in other endemic settings, possibly beyond dengue. Generalisability stretches past the models themselves, including the methods and procedures involved in i) the model development (exploratory analyses, model training, time series cross-validation, information criteria and scoring rules, and model comparison), ii) the reproducible computational pipeline (from processing heterogeneous data streams to iteratively updating forecasts and evaluations), and iii) the public-health-focused outputs (outbreak and surge detection indicators).

It is important for us to discuss the conflict of developing generative AI models to forecast epidemics of climate-sensitive infectious diseases. As noted in our perspective [53] in Chapter 9, the environmental costs of training data-hungry AI models for epidemics are often overlooked. Generative AI models have detrimental environmental impacts which are getting worse with rising energy and water needs for such models [415,416,752]. These effects are partially balanced against i) the aspired public health utility of our framework, ii) our extensive validation of AI models against epidemiologically naive baseline and traditional statistical models (to guard against novelty for the sake of novelty), and iii) our usage of pre-trained foundational time series models which reduce computational overload. However, these arguments are insufficient on their own and it would be instructive to further explore efficiency improvements in our pipeline in Chapter 4. As noted in Bashir et al. [752], efficiency improvements alone cannot promote sustainable practices. In future work, we should embrace recommendations to i) monitor the environmental impacts of pre-training, training, and/or fine-tuning models and ii) adopt multidisciplinary approaches to cost-benefit analyses [415,416,752], likely using our evaluation framework of Chapter 8 to quantify the value of using different models to inform decision-making.

Implementation of a decision support forecasting tool

Another current limitation is the ability to translate our models into real-time forecasting of dengue epidemic trajectories to routinely support public health authorities. This operationalisation will be challenging but will be aided by i) our previous work (Chapter 4), where we constructed and validated models using conditions that mimic real-world data availability, limited computational requirements, and developed a reproducible pipeline, and ii) our ongoing collaborations with public health authorities and researchers in Peru, Panama, Chile, Brazil, Thailand, Vietnam and French Polynesia.

Our aim here is to address a gap in the landscape of automated digital prediction tools for climate-sensitive infectious diseases [38] – tools created in partnership with those in affected, resource-limited settings. Our first step for the deployment of a modular framework is our collaboration with the Peru CDC, relying on the same data sources and pipeline as Chapter 4, with a view to then extending to the other countries and longer forecast horizons. The feasibility of scaling up our approach, as noted by Ryan et al. [279], will rest on robust model validation and software engineering practices, which we aim to achieve with ongoing extensions for the computational pipeline, open-source code to encourage collaborative efforts, and future works.

To promote incorporation in the decision-making procedures of public health authorities, we could use guidance from another field of science, implementation science [753]. This field considers how best to translate what we know (research and

evidence) into practice (public health policy). Implementation science has been promoted to optimise the implementation of tools in public health such as climate change threats[697], and could allow us to better understand and overcome obstacles to implementation. Meanwhile, for climate-sensitive infectious diseases, a 3-U (useful, usable, and used) framework has been proposed for digital prediction tools[280]. By first addressing the needs of the 3-U framework, implementation science could make adoption and integration more likely.

To date, we promote framework usefulness/reliability with multi-model approaches developed using extensive statistical theory (e.g., uncertainty quantification) and robust software engineering practices (e.g., containerisation). However, the forecast evaluations of Chapter 4 could be improved by our recent user-focused evaluation framework (Chapter 8). Alongside the evaluation summaries of forecast reliability, we could enhance trust in our models and forecasts by showing why past epidemics occurred and how these findings inform modelling choices. To ensure greater usefulness for decision-makers, we would need to extend forecast horizons to enable mid-range and seasonal forecasts, possibly using historical and forecasted weather data (e.g., using new deep learning forecasting models[286]) and predictions of serotype circulation (e.g.,[754]). In this thesis, we have not developed models for scenario predictions. Developing these models could be aided by our theory for retrospective, real-time, and forecasting analyses. We suspect that using scenario predictions will be important when there is a likely modifiable future (see below). In general, the epidemic dynamics are forecasted and usually not any changes to public health policy. An example where scenario predictions may be crucial would be likely policy changes to interventions such as vector control strategies, possibly influenced by outputs of operational models.

For usability, our model outputs balance statistical rigour and public health focus, while we emphasised containerisation and modularity throughout our computational pipeline. Therefore, once operational, researchers will be able to use different data sources, implement different forecasting models, and assess differences in forecast performance. We adopted lessons from the few efforts for deploying operational forecasting tools (e.g., visualisation for D-MOSS in Vietnam, Malaysia, and Sri Lanka[162]). By working closely with those who collect data and are responsible for control strategies in our studied regions, we aim to ensure that we, as modellers, understand the data such that forecasts match the priorities of decision-makers. The statistical and health literacy of a decision-maker, alongside a modeller's knowledge of local factors and data, are assumed here. These assumptions are non-trivial and also pertains to modelling both the past and present epidemic dynamics. Communicating uncertainty in model-based estimates about a highly nonlinear epidemic process is vital to reliably inform public health policy (e.g., preventing artificially high confidence)[103,435] but difficult[18].

For forecasting, there are additional complexities in such communication, in simple language, about a future that is inherently uncertain and often modifiable[15,158]. Although there is little scientific agreement on uncertainty and the language of probability[18,755], probability yardsticks are used in UK and US intelligence analysis[756] and could succinctly communicate the degree of uncertainty contained in the forecast for the decision-maker's decision problem. Another route to improve usability would be a dashboard with interpretable visualisations of predictions, such as traffic light schemes[168,540], warning impact matrices from weather forecasting[679], and/or visual summaries of historical forecast reliability. As we will discuss in Section 10.3.2, all these communication and literacy concerns form a critical outstanding issue that must be prioritised to ensure that our forecasting pipeline maintains clarity of audience and meets the needs of decision-makers and the public. Finally, we can further promote usability and effectiveness via our new user-focused evaluation framework of Chapter 8 and the associated, regular user surveys (e.g., to understand how best to communicate results to different stakeholders).

Reported data, causality, and mechanistic modelling

Forecasting reported dengue cases for a pathogen with high asymptomatic proportions and low reporting rates[7,178,179] is imperfect, compared to ideal latent DENV infections. The lack of age and serotype stratification in case data also prevents us from providing more specific forecasts that inform more targeted public health policy (e.g., demographic-specific vaccination strategies). These are limitations of the data, yet could constrain the utility of an operational forecasting

framework for supporting decision-making to reduce the morbidity and mortality of dengue. However, producing forecasts for observed data remains the most rigorous way that we can probabilistically evaluate the validity of our models' predictions (see below).

In the absence of complementary data streams, stratifications, or alternative epidemic outcomes such as hospitalisations, we tried to address the reporting and serotype limitations in our Bayesian hierarchical models (Chapter 4) by incorporating time- and space-varying random effects. In addition, we performed exploratory and sensitivity analyses and incorporated local knowledge. Based on evidence from Cambodia[438] and Thailand[236], the circulating serotypes, population immunity levels, the age distribution of infections, and reporting rates may not vary drastically over the short-term horizons that our models target. Invoking the Big Data Paradox of Meng[249], there is no guarantee that additional data would improve forecast performance. To promote relevant and sustainable data collection, VoI in simulation and empirical analyses[700] could again quantify the settings when and where information gains are largest for improving forecast reliability and informing related decisions.

However, we applied our forecasting framework (Chapter 4) to data from the northern coast of Peru – an area with endemic transmission and known circulation of two serotypes. Using seroprevalence and age-stratified case data, it would be possible to directly model the susceptible population and serotype-specific transmission patterns. This could address i) potential nonlinearities in epidemic curves and data distribution shifts induced by the build-up of population immunity[220] and DENV serotype competition[574]), and ii) the likely variations in reporting rates across space and time. This potentially time- and space-varying mechanistic structure to transmission is not directly modelled in our Bayesian hierarchical models. In addition to our usage of reported case data, these factors may partially explain why these models underperform, relative to other models, in higher spatial resolution. Poorer predictive performance could, but does not necessarily, limit the validity of causal conclusions from retrospective analyses.

This underperformance of a semi-mechanistic model for a short-term forecast horizon (one month) is consistent with previous similar forecasting studies for COVID-19[223] and dengue[169]. This may seem counterintuitive as we equipped a model with an estimated understanding of historical influences on epidemic dynamics, yet recall that we predicted the reported *case* data and not true unobserved *infections*. There is a difference between explaining the true infection counts and predicting the observed case counts[757]. The confounding effects of the aforementioned, unmodelled phenomena (e.g., unknown serotypes, population immunity, and reporting rates) and other influences (e.g., human mobility) may be less well represented by a parametric model fitted to all case data. For example, recent high precipitation may have different effects on reported dengue case incidence for different serotypes, for different levels of population immunity, for different human behaviour, and for different reporting rates. On the other hand, flexible deep learning and statistical time series models can indirectly handle non-stationarity and complex relationships such as data distribution shifts, while leveraging the predictable epidemic patterns such as seasonality or autoregressive behaviour. In future work, we could consider how this relative performance varies for longer horizons and/or for scenario predictions, where we suspect that having "correct" or plausible mechanistic structures will be vital for reliable forecasts, and more flexible, data-driven models may suffer a lack of generalisability in the absence of mechanistic structures. Hybrid AI approaches, integrating mechanisms in deep learning forecasting models, could be an interesting future direction to exploit the complementary benefits of realism and flexibility.

As an alternative forecast target, based on dialogue with local authorities and collaborators in Brazil and French Polynesia, it is useful to have forecasts of hospital admission rates and/or ICU capacity. Hospitalisation data could also help to correct for delays in case reporting. Therefore, we advocate for more widespread, real-time availability of dengue-related hospitalisation data (e.g., laboratory-confirmed diagnoses of DENV) alongside corresponding demographic data (e.g., hospital catchment population sizes). Incident hospitalisations have become a new forecasting target for seasonal influenza forecasting in both the US[167] and the UK[633]. Integrating yet another data source into our framework would be

computationally simple due to our existing pipeline, and we could use similar forecasting models alongside a new functional mapping from reported cases to hospitalisations and/or ICU capacity. This mapping would likely be based on risk factors, such as age. In terms of forecast evaluation (Chapter 8), the available action set and preferences of decision-makers could vary with these alternative epidemic measures.

For mechanistic extensions, conditional on real-time data availability, we could incorporate estimates of $R(t)$ and the generation time distribution using the framework and formulae of Chapter 6. Working with researchers in French Polynesia, we are developing a prototype temperature-dependent forecasting model that embeds these ideas into a renewal equation framework, alongside population immunity estimates and weather forecast data. This should add greater realism to forecasts, as $R(t)$ controls the real-time transmissibility, the generation time distribution controls the times between infections, and population immunity levels control the shapes of annual epidemic curves.

Our ensemble framework and reproducible pipeline can easily incorporate any predictive model, including scenario-based models, regardless of class/discipline (statistical, mechanistic, and/or AI-based, Figure 2.3). Using our forecast evaluation methods of Chapter 8, we can measure predictive improvements for decision-making when including different epidemic drivers within individual or different frameworks. This could improve the reliability and utility of our forecasting models for public health policy and the robustness of causal conclusions (Section 10.1.1). Alternative predictive causal options include counterfactual predictions where we permute different variables, for example, temperature or precipitation, and measure the effects on our models' predictions.

Extending the evidence synthesis

We could also extend our new explainability formulae from Chapter 5 to estimate the relative importance of different epidemic features/covariates at different forecast horizons. These findings can be compared with model-based insights from both retrospective and real-time analyses. This is important as without the understanding of *how* or *why* transmission has occurred, forecasts or scenario predictions are limited in their ability to inform effective control strategies.

In some ways, aspects of this idea are not new. It reflects the (disputed[758,759]) story of John Snow's discovery of water-borne transmission of cholera, as *description* and *data analysis* can lead to *explanation* to be used for both *prediction* and *prevention*[116].

In this thesis, we extend this integrated approach by working in all directions (see Figure 10.1). We argue that the insights of predictive forecasting models should inform (semi)-mechanistic models of the past and present, and vice versa. Extracting such theoretical and practical insights from different perspectives should support our overall objectives of i) better understanding of the past, present, and likely future dengue transmission and ii) more reliable and useful policy-focused outputs.

Combining a mechanistic understanding of a biological system with a data-driven understanding from prediction is a proposition that can be transferred to other biological systems. Insufficient or noisy data can prevent a direct mechanistic machine learning approach for biological systems[348], but testing the conclusions of our mechanistic model in predictive data-driven models can add robustness to such conclusions. As a mechanistic model is causal by design, our approach is also advantageous because inferring causality from observed data alone (i.e., from data-driven forecasting) is flawed and potentially misleading. This integrated and almost model-class-agnostic perspective has been adopted throughout the thesis (e.g., Chapters 4 and 5) and can be expanded in future works.

10.3.2 Forecast evaluation

An evaluation framework for actionable forecasting

As forecasts of epidemic trajectories can directly impact important policy decisions, they can influence individuals' health outcomes, quality of life, and broader societal well-being. It is therefore vital that forecasts are timely, reliable and useful for decision-makers. Recent research advances have focused forecast evaluation on alignment of probabilistic predictions with observations. These advances use scoring rules that measure the whole predictive distribution. Although a 'good' or accurate forecast depends on such alignment, the forecast is only useful (i.e., it has *value*) if it informs the decision-maker's dilemma; to take action or not.

Leaning on advances in infectious disease modelling and other branches of science (particularly weather forecasting, decision theory, and information theory), we developed an evaluation framework in Chapter 8 for forecasters and forecast evaluators. The framework emphasises forecast value and actionability for specific decision-makers and their specific policy/research questions, priorities, and associated risk preferences. In this work, we i) reinterpreted widely used scoring metrics of model performance as interpretable and public-health relevant quantities, ii) employed user- and event-focused metrics that enhance actionability by realigning evaluation with decision-makers' priorities and iii) developed theoretical foundations for risk-averse decision-making and the relationship between predictability of the epidemic and forecast performance. In short, this framework refocuses forecast evaluation on the decision-maker and simultaneously balances statistical rigour and public health priorities for the goal of more useful infectious disease forecasts.

We adopted ideas from other fields, but we also recognised the inherent differences in these fields (e.g., with weather forecasting[288]), and therefore carefully translated metrics and ideas into our epidemic context. We detailed crucial considerations for high-stakes decisions, and illustrated the application of a workflow using forecasts of state-level COVID-19 cases in the US. We found that ensemble models often provided the most clear value over individual models for different decision-makers, epidemic events, and across space and time, particularly during times of greater unpredictability. However, focusing on a specific wave of COVID-19 in a single state, we captured how the selected models varied depending on the policy/research question, the decision-maker's risk preferences, and over time. This highlighted i) the lack of any one-size-fits-all approach to evaluation and ii) the need for careful application, across space and time, of our evaluation framework to ensure model selection adapts to transient and uncertain decision-maker priorities and evolving epidemic dynamics.

As discussed in Chapter 8, future extensions include: i) embedding explicit utility functions[160] which will require eliciting decision-maker preferences for different policy/research questions in real time, and ii) developing additional methods for predictability analyses such as predicting future epidemic predictability as a function of past predictability and extending to multivariate permutation entropy[688] (as epidemic predictability may be greater conditional on other data).

From statistical theory to public health policy

This statistical theory has practical implications for future epidemic forecasting in terms of model development, forecast evaluation, and consequent public health policy. First, the metrics enable forecasters and evaluators to diagnose the statistical strengths and limitations of models through the lens of specific policy/research questions, specific epidemic outcomes/forecast targets, and specific decision-maker's priorities and risk preferences. This should encourage public-health-focused improvements to models and the reporting of reliable quantities by determining *a priori* whether a given model's forecast is likely to be reliable. Relatedly, the metrics are based on an "optimal" decision-making framework within the confines of a specific decision-maker, defined epidemic event, and set of models. This should promote better and more informed detections, preventions, and responses. Predictability analyses can complement these metric-based evaluations by guiding model selection and ensemble weighting, guarding against data distribution shifts and uncertainty,

and informing data collection.

The evaluation framework should empower local decision-makers and protect against erosion of public trust. This is because public health planning decisions can be made locally. For example, decisions about providing extra ICU capacity can be informed by the best available current information about the epidemic, the set of models, and the decision-maker's risk preferences and priorities. Although there will never be a single one-size-fits-all or optimal approach to epidemic decision-making or forecasting[160], a pre-defined, systematic evaluation framework provides defensibility of forecast-informed decisions that must be made under considerable uncertainty and time pressure.

Although the evaluation framework measures forecast value for decision-making, this should not be conflated with the use of forecasts as the sole informants to complex and uncertain decision-making. Recognising that forecasting models are just one part of the multidisciplinary decision-making process (see Figure 2.2 and[160]), this framework should be paired with outputs from other branches of science such as social science and economics. As modellers, we could also pair the results of an evaluation framework with scenario predictions[158] and with iterative updates to decision-maker risk preferences.

Local decision-making and uncertainty

This leads us to an important question for both our dengue forecast pipeline and our forecast evaluation workflow (Figures 4.7 and 8.1) about how local decision-makers can use the outputs of these frameworks in practice.

As suggested in the 3-U recommendations[280], a framework is only valuable if used. Forecasting models[3] and interventions[699] are sometimes developed without addressing the operationally relevant priorities of local decision-makers. Our evaluation workflow for forecasters and forecast evaluators starts with co-defining the policy/research questions with decision-makers, from which epidemic quantities are selected for forecasting and evaluation. The key outstanding next steps for evaluation are i) defining relevant questions and measured epidemic outcomes/events, and ii) eliciting decision-maker preferences in real time for the defined epidemic event and specific action. Both steps must be performed under considerable time pressures and uncertainty, while balancing ethical dilemmas and economic realities.

Co-defining the questions and epidemic events again assumes health and statistical literacy for decision-makers or scientific advisors in affected regions[105], and familiarity of the forecaster with the data, feasible actions, and available resources. More generally, health and statistical literacy are vital for the success of any quantitative approach that aims to empower informed local decision-making in outbreaks. Examples which promote this empowerment include knowledge sharing programmes paired with ongoing communication and support after a training workshop[760]. As identified in Chapter 1, there are important lessons from the inequalities of the COVID-19 pandemic and the history of infectious diseases. Funding should be prioritised for affected regions that have less experience or formal training in using data for managing outbreaks[24,699,761]. We hope that our various works encourage such funding. For example, we aimed for an evaluation framework that empowers decision-makers, focuses on their priorities, and promotes better public health responses.

Eliciting the decision-maker's degree of risk aversion via C/L ratios is non-trivial. Indeed, it could be more challenging than defining the relevant question and outcome/event. This ratio will never be known in real time with certainty. In Chapter 8, we described how these ratios could be elicited, but our applications recognised that the utility function of a decision-maker (via C/L ratios) will likely be ill-specified and uncertain. We therefore evaluated the forecast value across bands of uncertain preferences. To obtain reasonable ranges of such ratios for decision-makers, we propose accessible surveys of different decision-makers (similar to survey studies in weather forecasting[681]). The survey regularity will likely be determined by the evolving epidemic dynamics and associated transience of a decision-maker's C/L ratios. The transience is likely, as a decision-maker's risk preferences for the 'same' event (e.g., incidence above threshold) will differ in different settings, including different historical transmission patterns, available resources, underlying populations (demographics, immunity profile, and population health), and many other factors.

Once again, communication and collaboration will be central to all evaluations. The way we communicate scientific uncertainty has been found to have asymmetric effects on how people react; with people reacting more strongly when quality of evidence cues are low[678]. Recognising the aforementioned need to ensure statistical literacy among stakeholders, close and multilingual collaboration will be vital, alongside presentation of results with explainers (e.g., icon arrays)[105] and simple yet clear visualisations of uncertainty[103,203].

Finally, we recognise again that forecasts are just one strand of the decision-maker's toolkit. Framework applications should be complemented by i) performing further modelling tasks such as retrospective analyses of past dynamics, real-time estimates of $R(t)$, or scenario predictions, and ii) involving health economists and social scientists who can support the elicitation of decision-makers' priorities and risk preferences (e.g., economic evaluations[680]) and communication of evidence (e.g., lessons from psychology to supply succinct information[762]).

10.4 Conclusion

A model is just one way to describe the highly nonlinear and stochastic systems of epidemic spread, a data stream is just one way to parameterise the model, and both are only relevant and of "high-quality" for a specific, well-defined policy/research question. These models can target different questions, yet their outputs are not mutually exclusive, and we propose an integrated approach that stretches across models of the past, present, and future to unveil the evolving picture of complex epidemic dynamics.

Integrating heterogeneous data sources and comparing where different approaches agree and disagree are vital for science, as we try to better understand and communicate the uncertainty, value, and limits of models and strive towards more informed public health policy decisions. In this thesis, we recognise that our ultimate success in epidemiological modelling rests on i) intensive data collection efforts, ii) robust theory and model development that form the foundations for reliably answering public-health-relevant questions, iii) user-focused synthesis and communication of results, and iv) collaboration with local experts and public health authorities. These are key to the global health equity and decisions with positive public health impacts that our models must aim for.

We hope that the lessons learned from our model theory, evaluation frameworks, analyses, and overall evidence synthesis approach, from across branches of science, will advance scientific discovery, enhance pandemic preparedness, and inform public health responses in resource-limited settings.

Bibliography

[1] Anthony Fauci. A Goal of Service to Humankind. *NPR.org*, July 2005. URL https://www.npr.org/2005/07/25/4761448/a-goal-of-service-to-humankind. Publisher: NPR.org.

[2] Richard P. Feynman. *The character of physical law*. Modern Library, New York, modern library ed edition, 1965. ISBN 978-0-679-60127-2.

[3] Oliver Morgan. How decision makers can use quantitative approaches to guide outbreak responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 374(1776):20180365, July 2019. ISSN 1471-2970. doi: 10.1098/rstb.2018.0365.

[4] Christophe Fraser, Steven Riley, Roy M. Anderson, and Neil M. Ferguson. Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16): 6146–6151, April 2004. ISSN 0027-8424. doi: 10.1073/pnas.0307506101.

[5] Moritz U. G. Kraemer, Oliver G. Pybus, Christophe Fraser, Simon Cauchemez, Andrew Rambaut, and Benjamin J. Cowling. Monitoring key epidemiological parameters of SARS-CoV-2 transmission. *Nature Medicine*, 27(11): 1854–1855, November 2021. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-021-01545-w. URL https://www.nature.com/articles/s41591-021-01545-w.

[6] Azizur Rahman, Md Abdul Kuddus, Ryan H. L. Ip, and Michael Bewong. A Review of COVID-19 Modelling Strategies in Three Countries to Develop a Research Framework for Regional Areas. *Viruses*, 13(11):2185, October 2021. ISSN 1999-4915. doi: 10.3390/v13112185. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8623457/.

[7] João Bosco Siqueira Junior, Eduardo Massad, Abner Lobao-Neto, Randee Kastner, Louisa Oliver, and Elaine Gallagher. Epidemiology and costs of dengue in Brazil: a systematic literature review. *International Journal of Infectious Diseases*, 122:521–528, September 2022. ISSN 12019712. doi: 10.1016/j.ijid.2022.06.050. URL https://linkinghub.elsevier.com/retrieve/pii/S1201971222003836.

[8] World Health Organization. WHO Coronavirus (COVID-19) Dashboard, June 2025. URL https://covid19.who.int/.

[9] Rory C. O'Connor, Karen Wetherall, Seonaid Cleare, Heather McClelland, Ambrose J. Melson, Claire L. Niedzwiedz, Ronan E. O'Carroll, Daryl B. O'Connor, Steve Platt, Elizabeth Scowcroft, Billy Watson, Tiago Zortea, Eamonn Ferguson, and Kathryn A. Robb. Mental health and well-being during the COVID-19 pandemic: longitudinal analyses of adults in the UK COVID-19 Mental Health & Wellbeing study. *The British Journal of Psychiatry*, 218(6):326–333, June 2021. ISSN 0007-1250, 1472-1465. doi: 10.1192/bjp.2020.212. URL https://www.cambridge.org/core/product/identifier/S0007125020002123/type/journal_article.

[10] A. Sud, M.E. Jones, J. Broggio, C. Loveday, B. Torr, A. Garrett, D.L. Nicol, S. Jhanji, S.A. Boyce, F. Gronthoud, P. Ward, J.M. Handy, N. Yousaf, J. Larkin, Y-E. Suh, S. Scott, P.D.P. Pharoah, C. Swanton, C. Abbosh, M. Williams,

G. Lyratzopoulos, R. Houlston, and C. Turnbull. Collateral damage: the impact on outcomes from cancer surgery of the COVID-19 pandemic. *Annals of Oncology*, 31(8):1065–1074, August 2020. ISSN 09237534. doi: 10.1016/j.annonc.2020.05.009. URL https://linkinghub.elsevier.com/retrieve/pii/S0923753420398252.

[11] Robin Y. Park, Alyssa Bilinski, Robbie M. Parks, and Seth Flaxman. Trends in Maternal, Fetal, and Infant Mortality in the US, 2000-2023. *JAMA Pediatrics*, April 2025. ISSN 2168-6203. doi: 10.1001/jamapediatrics.2025.0440. URL https://jamanetwork.com/journals/jamapediatrics/fullarticle/2833316.

[12] Nicolas Wentzensen, Megan A. Clarke, and Rebecca B. Perkins. Impact of COVID-19 on cervical cancer screening: Challenges and opportunities to improving resilience and reduce disparities. *Preventive Medicine*, 151:106596, October 2021. ISSN 00917435. doi: 10.1016/j.ypmed.2021.106596. URL https://linkinghub.elsevier.com/retrieve/pii/S0091743521001808.

[13] Christina J. Atchison, Bethan Davies, Emily Cooper, Adam Lound, Matthew Whitaker, Adam Hampshire, Adriana Azor, Christl A. Donnelly, Marc Chadeau-Hyam, Graham S. Cooke, Helen Ward, and Paul Elliott. Long-term health impacts of COVID-19 among 242,712 adults in England. *Nature Communications*, 14(1):6588, October 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-41879-2. URL https://www.nature.com/articles/s41467-023-41879-2.

[14] Hannah E. Davis, Gina S. Assaf, Lisa McCorkell, Hannah Wei, Ryan J. Low, Yochai Re'em, Signe Redfield, Jared P. Austin, and Athena Akrami. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *eClinicalMedicine*, 38:101019, August 2021. ISSN 25895370. doi: 10.1016/j.eclinm.2021.101019. URL https://linkinghub.elsevier.com/retrieve/pii/S2589537021002996.

[15] Anne Cori and Adam Kucharski. Inference of epidemic dynamics in the COVID-19 era and beyond. *Epidemics*, 48:100784, September 2024. ISSN 17554365. doi: 10.1016/j.epidem.2024.100784. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436524000458.

[16] Ruth McCabe and Christl A. Donnelly. Disease transmission and control modelling at the science–policy interface. *Interface Focus*, 11(6):20210013, December 2021. ISSN 2042-8901. doi: 10.1098/rsfs.2021.0013. URL https://royalsocietypublishing.org/doi/10.1098/rsfs.2021.0013.

[17] Ellen Brooks-Pollock, Leon Danon, Thibaut Jombart, and Lorenzo Pellis. Modelling that shaped the early COVID-19 pandemic response in the UK. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829): 20210001, May 2021. doi: 10.1098/rstb.2021.0001. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2021.0001. Publisher: Royal Society.

[18] Graham F. Medley. A consensus of evidence: The role of SPI-M-O in the UK COVID-19 response. *Advances in Biological Regulation*, 86:100918, December 2022. ISSN 2212-4926. doi: 10.1016/j.jbior.2022.100918. URL https://www.sciencedirect.com/science/article/pii/S2212492622000586.

[19] Sangeeta Bhatia, Natsuko Imai, Oliver J Watson, Auss Abbood, Philip Abdelmalik, Thijs Cornelissen, Stéphane Ghozzi, Britta Lassmann, Radhika Nagesh, Manon L Ragonnet-Cronin, Johannes Christof Schnitzler, Moritz Ug Kraemer, Simon Cauchemez, Pierre Nouvellet, and Anne Cori. Lessons from COVID-19 for rescalable data collection. *The Lancet Infectious Diseases*, 23(9):e383–e388, September 2023. ISSN 14733099. doi: 10.1016/S1473-3099(23)00121-4. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309923001214.

[20] Caroline Buckee. Improving epidemic surveillance and response: big data is dead, long live big data. *The Lancet Digital Health*, 2(5):e218–e220, May 2020. ISSN 25897500. doi: 10.1016/S2589-7500(20)30059-5. URL https://linkinghub.elsevier.com/retrieve/pii/S2589750020300595.

[21] Lucy Goodfellow, Edwin Van Leeuwen, and Rosalind M. Eggo. COVID-19 inequalities in England: a mathematical modelling study of transmission risk and clinical vulnerability by socioeconomic status. *BMC Medicine*, 22(1):162, April 2024. ISSN 1741-7015. doi: 10.1186/s12916-024-03387-y. URL https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-024-03387-y.

[22] Rachid Laajaj, Duncan Webb, Danilo Aristizabal, Eduardo Behrentz, Raquel Bernal, Giancarlo Buitrago, Zulma Cucunubá, Fernando de la Hoz, Alejandro Gaviria, Luis Jorge Hernández, Camilo De Los Rios, Andrea Ramírez Varela, Silvia Restrepo, Norbert Schady, and Martha Vives. Understanding how socioeconomic inequalities drive inequalities in COVID-19 infections. *Scientific Reports*, 12(1):8269, May 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-11706-7. URL https://www.nature.com/articles/s41598-022-11706-7. Publisher: Nature Publishing Group.

[23] Richard Blundell, Monica Costa Dias, Jonathan Cribb, Robert Joyce, Tom Waters, Thomas Wernham, and Xiaowei Xu. Inequality and the COVID-19 Crisis in the United Kingdom. *Annual Review of Economics*, 14(Volume 14, 2022): 607–636, August 2022. ISSN 1941-1383, 1941-1391. doi: 10.1146/annurev-economics-051520-030252. URL https://www.annualreviews.org/content/journals/10.1146/annurev-economics-051520-030252. Publisher: Annual Reviews.

[24] Jeffrey D Sachs, Salim S Abdool Karim, Lara Aknin, Joseph Allen, Kirsten Brosbøl, Francesca Colombo, Gabriela Cuevas Barron, María Fernanda Espinosa, Vitor Gaspar, Alejandro Gaviria, Andy Haines, Peter J Hotez, Phoebe Koundouri, Felipe Larraín Bascuñán, Jong-Koo Lee, Muhammad Ali Pate, Gabriela Ramos, K Srinath Reddy, Ismail Serageldin, John Thwaites, Vaira Vike-Freiberga, Chen Wang, Miriam Khamadi Were, Lan Xue, Chandrika Bahadur, Maria Elena Bottazzi, Chris Bullen, George Laryea-Adjei, Yanis Ben Amor, Ozge Karadag, Guillaume Lafortune, Emma Torres, Lauren Barredo, Juliana G E Bartels, Neena Joshi, Margaret Hellard, Uyen Kim Huynh, Shweta Khandelwal, Jeffrey V Lazarus, and Susan Michie. The Lancet Commission on lessons for the future from the COVID-19 pandemic. *The Lancet*, 400(10359):1224–1280, October 2022. ISSN 01406736. doi: 10.1016/S0140-6736(22)01585-9. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673622015859.

[25] Sumali Bajaj, Siyu Chen, Richard Creswell, Reshania Naidoo, Joseph L-H Tsui, Olumide Kolade, George Nicholson, Brieuc Lehmann, James A Hay, Moritz U G Kraemer, Ricardo Aguas, Christl A Donnelly, Tom Fowler, Susan Hopkins, Liberty Cantrell, Prabin Dahal, Lisa J White, Kasia Stepniewska, Merryn Voysey, Ben Lambert, Ricardo Aguas, Ma'ayan Amswych, Billie Andersen-Waine, Sumali Bajaj, Kweku Bimpong, Adam Bodley, Liberty Cantrell, Siyu Chen, Richard Creswell, Prabin Dahal, Sophie Dickinson, Sabine Dittrich, Tracy Evans, Angus Ferguson-Lewis, Caroline Franco, Bo Gao, Rachel Hounsell, Muhammad Kasim, Claire Keene, Ben Lambert, Umar Mahmood, Melinda Mills, Ainura Moldokmatova, Sassy Molyneux, Reshania Naidoo, Randolph Ngwafor Anye, Jared Norman, Wirichada Pan-Ngum, Sunil Pokharel, Anastasiia Polner, Emily Rowe, Sompob Saralamba, Rima Shretta, Sheetal Silal, Kasia Stepniewska, Joseph L-H Tsui, Merryn Voysey, Marta Wanat, and Lisa J White. COVID-19 testing and reporting behaviours in England across different sociodemographic groups: a population-based study using testing data and data from community prevalence surveillance surveys. *The Lancet Digital Health*, 6(11):e778–e790, November 2024. ISSN 2589-7500. doi: 10.1016/s2589-7500(24)00169-9. URL https://linkinghub.elsevier.com/retrieve/pii/S2589750024001699. Publisher: Elsevier BV.

[26] Victoria J McGowan and Clare Bambra. COVID-19 mortality and deprivation: pandemic, syndemic, and endemic health inequalities. *The Lancet Public Health*, 7(11):e966–e975, November 2022. ISSN 2468-2667. doi: 10.1016/s2468-2667(22)00223-7. URL https://linkinghub.elsevier.com/retrieve/pii/S2468266722002237. Publisher: Elsevier BV.

[27] Christian Morgenstern, Thomas Rawson, Wes Hinsley, Pablo N. Perez Guzman, Samir Bhatt, and Neil M. Ferguson. Socioeconomic and temporal heterogeneity in SARS-CoV-2 exposure and disease in England from May 2020 to February 2023. *Science Advances*, 11(21):eadu8678, May 2025. ISSN 2375-2548. doi: 10.1126/sciadv.adu8678. URL https://www.science.org/doi/10.1126/sciadv.adu8678.

[28] Anderson F. Brito, Elizaveta Semenova, Gytis Dudas, Gabriel W. Hassler, Chaney C. Kalinich, Moritz U. G. Kraemer, Joses Ho, Houriiyah Tegally, George Githinji, Charles N. Agoti, Lucy E. Matkin, Charles Whittaker, Benjamin P. Howden, Vitali Sintchenko, Neta S. Zuckerman, Orna Mor, Heather M. Blankenship, Tulio de Oliveira, Raymond T. P. Lin, Marilda Mendonça Siqueira, Paola Cristina Resende, Ana Tereza R. Vasconcelos, Fernando R. Spilki, Renato Santana Aguiar, Ivailo Alexiev, Ivan N. Ivanov, Ivva Philipova, Christine V. F. Carrington, Nikita S. D. Sahadeo, Ben Branda, Céline Gurry, Sebastian Maurer-Stroh, Dhamari Naidoo, Karin J. von Eije, Mark D. Perkins, Maria van Kerkhove, Sarah C. Hill, Ester C. Sabino, Oliver G. Pybus, Christopher Dye, Samir Bhatt, Seth Flaxman, Marc A. Suchard, Nathan D. Grubaugh, Guy Baele, and Nuno R. Faria. Global disparities in SARS-CoV-2 genomic surveillance. *Nature Communications*, 13(1):7003, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-33713-y. URL https://www.nature.com/articles/s41467-022-33713-y. Publisher: Nature Publishing Group.

[29] Rachel E. Baker, Ayesha S. Mahmud, Ian F. Miller, Malavika Rajeev, Fidisoa Rasambainarivo, Benjamin L. Rice, Saki Takahashi, Andrew J. Tatem, Caroline E. Wagner, Lin-Fa Wang, Amy Wesolowski, and C. Jessica E. Metcalf. Infectious disease in an era of global change. *Nature Reviews Microbiology*, 20(4):193–205, April 2022. ISSN 1740-1526, 1740-1534. doi: 10.1038/s41579-021-00639-z. URL https://www.nature.com/articles/s41579-021-00639-z. Publisher: Springer Science and Business Media LLC.

[30] Priya Venkatesan. WHO world malaria report 2024. *The Lancet Microbe*, 6(4):101073, April 2025. ISSN 26665247. doi: 10.1016/j.lanmic.2025.101073. URL https://linkinghub.elsevier.com/retrieve/pii/S2666524725000011.

[31] World Health Organization. Disease Outbreak News; Dengue – Global situation. Technical report, May 2024. URL https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON518.

[32] Sashini Payagala and Anton Pozniak. The global burden of HIV. *Clinics in Dermatology*, 42(2):119–127, March 2024. ISSN 0738-081X. doi: 10.1016/j.clindermatol.2024.02.001. URL https://www.sciencedirect.com/science/article/pii/S0738081X24000336.

[33] Stefan Litvinjenko, Olivia Magwood, Shishi Wu, and Xiaolin Wei. Burden of tuberculosis among vulnerable populations worldwide: an overview of systematic reviews. *The Lancet Infectious Diseases*, 23(12):1395–1407, December 2023. ISSN 14733099. doi: 10.1016/S1473-3099(23)00372-9. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309923003729.

[34] P Farmer. Social inequalities and emerging infectious diseases. *Emerging Infectious Diseases*, 2(4):259–269, 1996. ISSN 1080-6040. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2639930/.

[35] Sandra Crouse Quinn and Supriya Kumar. Health inequalities and infectious disease epidemics: a challenge for global health security. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, 12(5):263–273, 2014. ISSN 1557-850X. doi: 10.1089/bsp.2014.0032.

[36] Mohsen Naghavi, Tomislav Mestrovic, Authia Gray, Anna Gershberg Hayoon, Lucien R Swetschinski, Gisela Robles Aguilar, Nicole Davis Weaver, Kevin S Ikuta, Erin Chung, Eve E Wool, Chieh Han, Daniel T Araki, Samuel B Albertson, Rose Bender, Greg Bertolacci, Annie J Browne, Ben S Cooper, Matthew W Cunningham, Christiane Dolecek, Matthew Doxey, Susanna J Dunachie, Sama Ghoba, Georgina Haines-Woodhouse, Simon I Hay, Rebecca L

Hsu, Kenneth C Iregbu, Hmwe H Kyu, Jorge R Ledesma, Jianing Ma, Catrin E Moore, Jonathan F Mosser, Vincent Mougin, Pirouz Naghavi, Amanda Novotney, Victor Daniel Rosenthal, Benn Sartorius, Andy Stergachis, Christopher Troeger, Avina Vongpradith, Magdalene K Walters, Han Yong Wunrow, and Christopher Jl Murray. Global burden associated with 85 pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Infectious Diseases*, 24(8):868–895, August 2024. ISSN 1473-3099. doi: 10.1016/s1473-3099(24)00158-0. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309924001580. Publisher: Elsevier BV.

[37] Jan C. Semenza, Joacim Rocklöv, and Kristie L. Ebi. Climate Change and Cascading Risks from Infectious Disease. *Infectious Diseases and Therapy*, 11(4):1371–1390, August 2022. ISSN 2193-8229, 2193-6382. doi: 10.1007/s40121-022-00647-3. URL https://link.springer.com/10.1007/s40121-022-00647-3.

[38] Sadie J. Ryan, Colin J. Carlson, Erin A. Mordecai, and Leah R. Johnson. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS Neglected Tropical Diseases*, 13(3):e0007213, March 2019. ISSN 1935-2727. doi: 10.1371/journal.pntd.0007213. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6438455/.

[39] Jing Liu-Helmersson, Mikkel Quam, Annelies Wilder-Smith, Hans Stenlund, Kristie Ebi, Eduardo Massad, and Joacim Rocklöv. Climate Change and *Aedes* Vectors: 21st Century Projections for Dengue Transmission in Europe. *EBioMedicine*, 7:267–277, May 2016. ISSN 2352-3964. doi: 10.1016/j.ebiom.2016.03.046. URL https://www.sciencedirect.com/science/article/pii/S2352396416301335.

[40] Jie Hu, Benjamin P. Horton, Tsin Wen Yeo, Joseph J. Y. Sung, and Yim Hung Lam Steve. Mosquito and global dengue cases in a warming world. *BMJ Global Health*, 10(5), May 2025. ISSN 2059-7908. doi: 10.1136/bmjgh-2023-014688. URL https://gh.bmj.com/content/10/5/e014688. Publisher: BMJ Publishing Group Ltd.

[41] European Centre for Disease Prevention and Control. Twelve-month dengue virus disease case notification rate per 100 000 population, January–December 2024, February 2025. URL https://www.ecdc.europa.eu/en/publications-data/twelve-month-dengue-virus-disease-case-notification-rate-100-000-population.

[42] World Health Organization. World Health Organization: Global dengue surveillance dashboard, June 2025. URL https://worldhealthorg.shinyapps.io/dengue_global/.

[43] World Health Organization. Global arbovirus initiative: preparing for the next pandemic by tackling mosquito-borne viruses with epidemic and pandemic potential. Technical report, May 2024. URL https://www.who.int/publications/i/item/9789240088948.

[44] Cathal Mills, Marc Chadeau-Hyam, Paul Elliott, and Christl A Donnelly. The utility of wastewater surveillance for monitoring SARS-CoV-2 prevalence. *PNAS Nexus*, page 438, October 2024. ISSN 2752-6542. doi: 10.1093/pnasnexus/pgae438. URL https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae438/7811170.

[45] Cathal Mills and Christl A. Donnelly. Climate-based modelling and forecasting of dengue in three endemic departments of Peru. *PLOS Neglected Tropical Diseases*, 18(12):e0012596, December 2024. ISSN 1935-2735. doi: 10.1371/journal.pntd.0012596. URL https://dx.plos.org/10.1371/journal.pntd.0012596.

[46] Cathal Mills, Francesca Falconi-Agapito, Jean-Paul Carrera, César V. Munayco, Moritz U. G. Kraemer, and Christl A. Donnelly. Multi-model approach to understand and predict past and future dengue epidemic dynamics, June 2025. URL https://www.medrxiv.org/content/10.1101/2024.10.18.24315690v2. Pages: 2024.10.18.24315690.

[47] Cathal Mills, Geziel dos Santos de Sousa, Antonio Silva Lima Neto, Vasco Furtado, Sen Pei, Moritz U. G. Kraemer, and Christl A. Donnelly. The time- and space-varying roles of human mobility in shaping urban dengue epidemics, June 2025. URL https://www.medrxiv.org/content/10.1101/2025.06.03.25328731v1. Pages: 2025.06.03.25328731.

[48] Cathal Mills, Tarek Alrefae, William S. Hart, Moritz U. G. Kraemer, Kris V. Parag, Robin N. Thompson, Christl A. Donnelly, and Ben Lambert. Renewal equations for vector-borne diseases, September 2024. URL http://arxiv.org/abs/2409.18726. arXiv:2409.18726 [q-bio].

[49] Cathal Mills and Ben Lambert. Exact simulation of stochastic reaction processes with external variability. June 2025.

[50] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, December 1976. ISSN 00219991. doi: 10.1016/0021-9991(76)90041-3. URL https://linkinghub.elsevier.com/retrieve/pii/0021999176900413.

[51] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, December 1977. ISSN 0022-3654, 1541-5740. doi: 10.1021/j100540a008. URL https://pubs.acs.org/doi/abs/10.1021/j100540a008.

[52] Cathal Mills, Nicholas J. Irons, Joseph L.-H. Tsui, Luiz Max Carvalho, Claudia T. Codeco, Adam J. Kucharski, Oliver Ratmann, Ben Lambert, Christl A. Donnelly, and Moritz U. G. Kraemer. From metric to action: An evaluation framework to translate infectious disease forecasts into policy decisions. June 2025.

[53] Moritz U. G. Kraemer, Joseph L.-H. Tsui, Serina Y. Chang, Spyros Lytras, Mark P. Khurana, Samantha Vanderslott, Sumali Bajaj, Neil Scheidwasser, Jacob Liam Curran-Sebastian, Elizaveta Semenova, Mengyan Zhang, H. Juliette T. Unwin, Oliver J. Watson, Cathal Mills, Abhishek Dasgupta, Luca Ferretti, Samuel V. Scarpino, Etien Koua, Oliver Morgan, Houriiyah Tegally, Ulrich Paquet, Loukas Moutsianas, Christophe Fraser, Neil M. Ferguson, Eric J. Topol, David A. Duchêne, Tanja Stadler, Patricia Kingori, Michael J. Parker, Francesca Dominici, Nigel Shadbolt, Marc A. Suchard, Oliver Ratmann, Seth Flaxman, Edward C. Holmes, Manuel Gomez-Rodriguez, Bernhard Schölkopf, Christl A. Donnelly, Oliver G. Pybus, Simon Cauchemez, and Samir Bhatt. Artificial intelligence for modelling infectious disease epidemics. *Nature*, 638(8051):623–635, February 2025. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-08564-w. URL https://www.nature.com/articles/s41586-024-08564-w.

[54] Caitriona Murphy, Wey Wen Lim, Cathal Mills, Jessica Y. Wong, Dongxuan Chen, Yanmy Xie, Mingwei Li, Susan Gould, Hualei Xin, Justin K. Cheung, Samir Bhatt, Benjamin J. Cowling, and Christl A. Donnelly. Effectiveness of social distancing measures and lockdowns for reducing transmission of COVID-19 in non-healthcare, community-based settings. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381(2257):20230132, October 2023. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2023.0132. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2023.0132.

[55] Mehrdad Halaji, Mohammad Heiat, Niloofar Faraji, and Reza Ranjbar. Epidemiology of COVID-19: An updated review. *Journal of Research in Medical Sciences*, 26(1):82, 2021. ISSN 1735-1995. doi: 10.4103/jrms.JRMS_506_20. URL https://journals.lww.com/10.4103/jrms.JRMS_506_20.

[56] Fuqing Wu, Amy Xiao, Jianbo Zhang, Katya Moniz, Noriko Endo, Federica Armas, Richard Bonneau, Megan A. Brown, Mary Bushman, Peter R. Chai, Claire Duvallet, Timothy B. Erickson, Katelyn Foppe, Newsha Ghaeli, Xiaoqiong Gu, William P. Hanage, Katherine H. Huang, Wei Lin Lee, Mariana Matus, Kyle A. McElroy, Jonathan Nagler, Steven F. Rhode, Mauricio Santillana, Joshua A. Tucker, Stefan Wuertz, Shijie Zhao, Janelle Thompson, and Eric J. Alm. SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation

of new COVID-19 cases. *Science of The Total Environment*, 805:150121, January 2022. ISSN 00489697. doi: 10.1016/j.scitotenv.2021.150121. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969721051962.

[57] Youssef Miyah, Mohammed Benjelloun, Sanae Lairini, and Anissa Lahrichi. COVID-19 Impact on Public Health, Environment, Human Psychology, Global Socioeconomy, and Education. *TheScientificWorldJournal*, 2022:5578284, 2022. ISSN 1537-744X. doi: 10.1155/2022/5578284.

[58] Johannes Bracher, Evan L. Ray, Tilmann Gneiting, and Nicholas G. Reich. Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology*, 17(2):e1008618, February 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008618. URL https://dx.plos.org/10.1371/journal.pcbi.1008618.

[59] Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Imperial College COVID-19 Response Team, Pablo N. Perez-Guzman, Nora Schmit, Lucia Cilloni, Kylie E. C. Ainslie, Marc Baguelin, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Laura V. Cooper, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine L. Van Elsland, Richard G. FitzJohn, Katy A. M. Gaythorpe, Lily Geidelberg, Nicholas C. Grassly, William D. Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, Edward Knock, Daniel J. Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris V. Parag, Igor Siveroni, Hayley A. Thompson, Robert Verity, Erik Volz, Caroline E. Walters, Haowei Wang, Yuanrong Wang, Oliver J. Watson, Peter Winskill, Xiaoyue Xi, Patrick G. T. Walker, Azra C. Ghani, Christl A. Donnelly, Steven Riley, Michaela A. C. Vollmer, Neil M. Ferguson, Lucy C. Okell, and Samir Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, August 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2405-7. URL https://www.nature.com/articles/s41586-020-2405-7.

[60] Thomas Hale, Noam Angrist, Andrew J. Hale, Beatriz Kira, Saptarshi Majumdar, Anna Petherick, Toby Phillips, Devi Sridhar, Robin N. Thompson, Samuel Webster, and Yuxi Zhang. Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. *PLOS ONE*, 16(7):e0253116, July 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0253116. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0253116. Publisher: Public Library of Science.

[61] Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Open COVID-19 Data Working Group, Louis Du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490):493–497, May 2020. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abb4218. URL https://www.science.org/doi/10.1126/science.abb4218.

[62] Nicholas C. Grassly and Christophe Fraser. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, 6(6):477–487, June 2008. ISSN 1740-1526, 1740-1534. doi: 10.1038/nrmicro1845. URL https://www.nature.com/articles/nrmicro1845.

[63] Jane P. Messina, Oliver J. Brady, Thomas W. Scott, Chenting Zou, David M. Pigott, Kirsten A. Duda, Samir Bhatt, Leah Katzelnick, Rosalind E. Howes, Katherine E. Battle, Cameron P. Simmons, and Simon I. Hay. Global spread of dengue virus types: mapping the 70 year history. *Trends in Microbiology*, 22(3):138–146, March 2014. ISSN 0966-842X. doi: 10.1016/j.tim.2013.12.011. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3946041/.

[64] Verity Hill, Sara Cleemput, James Siqueira Pereira, Robert J. Gifford, Vagner Fonseca, Houriiyah Tegally, Anderson F. Brito, Gabriela Ribeiro, Vinicius Carius De Souza, Isabela Carvalho Brcko, Igor Santana Ribeiro, Iago Trezena Tavares

De Lima, Svetoslav Nanev Slavov, Sandra Coccuzzo Sampaio, Maria Carolina Elias, Vi Thuy Tran, Duong Thi Hue Kien, Tuyen Huynh, Sophie Yacoub, Idrissa Dieng, Richard Salvato, Gabriel Luz Wallau, Tatiana S. Gregianini, Fernanda M. S. Godinho, Chantal B. F. Vogels, Mallery I. Breban, Mariana Leguia, Suraj Jagtap, Rahul Roy, Chanditha Hapuarachchi, Gaspary Mwanyika, Marta Giovanetti, Luiz C. J. Alcantara, Nuno R. Faria, Christine V. F. Carrington, Kathryn A. Hanley, Edward C. Holmes, Wim Dumon, Alex Ranieri Jerônimo Lima, Tulio De Oliveira, and Nathan D. Grubaugh. A new lineage nomenclature to aid genomic surveillance of dengue virus. *PLOS Biology*, 22(9):e3002834, September 2024. ISSN 1545-7885. doi: 10.1371/journal.pbio.3002834. URL https://dx.plos.org/10.1371/journal.pbio.3002834. Publisher: Public Library of Science (PLoS).

[65] Duane J. Gubler. Dengue/dengue haemorrhagic fever: history and current status. *Novartis Foundation Symposium*, 277:3–16; discussion 16–22, 71–73, 251–253, 2006. ISSN 1528-2511. doi: 10.1002/0470058005.ch2.

[66] Susumu Hotta. Experimental Studies on Dengue: I. Isolation, Identification and Modification of the Virus. *The Journal of Infectious Diseases*, 90(1):1–9, 1952. ISSN 0022-1899. URL https://www.jstor.org/stable/30093992. Publisher: Oxford University Press.

[67] World Health Organization. Disease Outbreak News; Dengue – Global situation. Technical report, December 2023. URL https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498.

[68] Cameron P. Simmons, Jeremy J. Farrar, Nguyen van Vinh Chau, and Bridget Wills. Dengue. *New England Journal of Medicine*, 366(15):1423–1432, April 2012. ISSN 0028-4793. doi: 10.1056/NEJMra1110265. URL https://www.nejm.org/doi/full/10.1056/NEJMra1110265. Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMra1110265.

[69] Ahyoung Lim, Freya M. Shearer, Kara Sewalk, David M. Pigott, Joseph Clarke, Azhar Ghouse, Ciara Judge, Hyolim Kang, Jane P. Messina, Moritz U. G. Kraemer, Katy A. M. Gaythorpe, William M. de Souza, Elaine O. Nsoesie, Michael Celone, Nuno Faria, Sadie J. Ryan, Ingrid B. Rabe, Diana P. Rojas, Simon I. Hay, John S. Brownstein, Nick Golding, and Oliver J. Brady. The overlapping global distribution of dengue, chikungunya, Zika and yellow fever. *Nature Communications*, 16(1):3418, April 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-58609-5. URL https://www.nature.com/articles/s41467-025-58609-5. Publisher: Nature Publishing Group.

[70] Oliver J. Brady and Simon I. Hay. The Global Expansion of Dengue: How *Aedes aegypti* Mosquitoes Enabled the First Pandemic Arbovirus. *Annual Review of Entomology*, 65(1):191–208, January 2020. ISSN 0066-4170, 1545-4487. doi: 10.1146/annurev-ento-011019-024918. URL https://www.annualreviews.org/doi/10.1146/annurev-ento-011019-024918.

[71] Gabriela Paz-Bailey, Laura E Adams, Jacqueline Deen, Kathryn B Anderson, and Leah C Katzelnick. Dengue. *The Lancet*, 403(10427):667–682, February 2024. ISSN 0140-6736. doi: 10.1016/S0140-6736(23)02576-X. URL https://www.sciencedirect.com/science/article/pii/S014067362302576X.

[72] European Centre for Disease Prevention and Control. Dengue worldwide overview; Situation update, December 2023, December 2023. URL https://www.ecdc.europa.eu/en/dengue-monthly.

[73] Olga De Santis, Nicolas Bouscaren, and Antoine Flahault. Asymptomatic dengue infection rate: A systematic literature review. *Heliyon*, 9(9):e20069, September 2023. ISSN 2405-8440. doi: 10.1016/j.heliyon.2023.e20069. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10559824/.

[74] Najmul Haider, Mohammad Nayeem Hasan, Joshua Onyango, and Md Asaduzzaman. Global landmark: 2023 marks the worst year for dengue cases with millions infected and thousands of deaths reported. *IJID Regions*, 13:100459,

September 2024. ISSN 2772-7076. doi: 10.1016/j.ijregi.2024.100459. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11532885/.

[75] Andrew Teo, Hao Dong Tan, Thomas Loy, Po Ying Chia, and Caroline Lin Lin Chua. Understanding antibody-dependent enhancement in dengue: Are afucosylated IgG1s a concern? *PLOS Pathogens*, 19(3):e1011223, March 2023. ISSN 1553-7366. doi: 10.1371/journal.ppat.1011223. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10062565/.

[76] The Lancet. Dengue: the threat to health now and in the future. *The Lancet*, 404(10450):311, July 2024. ISSN 01406736. doi: 10.1016/S0140-6736(24)01542-3. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673624015423.

[77] Duane J. Gubler. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21st Century. *Tropical Medicine and Health*, 39(4SUPPLEMENT):S3–S11, 2011. ISSN 1349-4147, 1348-8945. doi: 10.2149/tmh.2011-S05. URL https://www.jstage.jst.go.jp/article/tmh/39/4SUPPLEMENT/39_2011-S05/_article.

[78] Rachel Lowe, Sophie A Lee, Kathleen M O'Reilly, Oliver J Brady, Leonardo Bastos, Gabriel Carrasco-Escobar, Rafael De Castro Catão, Felipe J Colón-González, Christovam Barcellos, Marilia Sá Carvalho, Marta Blangiardo, Håvard Rue, and Antonio Gasparrini. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *The Lancet Planetary Health*, 5(4):e209–e219, April 2021. ISSN 25425196. doi: 10.1016/S2542-5196(20)30292-8. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519620302928.

[79] Rory Gibb, Felipe J. Colón-González, Phan Trong Lan, Phan Thi Huong, Vu Sinh Nam, Vu Trong Duoc, Do Thai Hung, Nguyn Thanh Dong, Vien Chinh Chien, Ly Thi Thuy Trang, Do Kien Quoc, Tran Minh Hoa, Nguyen Hu Tai, Tran Thi Hang, Gina Tsarouchi, Eleanor Ainscoe, Quillon Harpham, Barbara Hofmann, Darren Lumbroso, Oliver J. Brady, and Rachel Lowe. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. *Nature Communications*, 14(1):8179, December 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-43954-0. URL https://www.nature.com/articles/s41467-023-43954-0.

[80] Miranda Chan and Michael A. Johansson. The incubation periods of dengue viruses. *PLOS ONE*, 7(11):e50972, November 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0050972. URL https://dx.plos.org/10.1371/journal.pone.0050972.

[81] Oliver J Brady, Michael A Johansson, Carlos A Guerra, Samir Bhatt, Nick Golding, David M Pigott, Hélène Delatte, Marta G Grech, Paul T Leisnham, Rafael Maciel-de Freitas, Linda M Styer, David L Smith, Thomas W Scott, Peter W Gething, and Simon I Hay. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasites & Vectors*, 6(1):351, December 2013. ISSN 1756-3305. doi: 10.1186/1756-3305-6-351. URL https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-6-351.

[82] Erin A. Mordecai, Jeremy M. Cohen, Michelle V. Evans, Prithvi Gudapati, Leah R. Johnson, Catherine A. Lippi, Kerri Miazgowicz, Courtney C. Murdock, Jason R. Rohr, Sadie J. Ryan, Van Savage, Marta S. Shocket, Anna Stewart Ibarra, Matthew B. Thomas, and Daniel P. Weikel. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLOS Neglected Tropical Diseases*, 11(4):e0005568, April 2017. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005568. URL https://dx.plos.org/10.1371/journal.pntd.0005568.

[83] Erin A. Mordecai, Jamie M. Caldwell, Marissa K. Grossman, Catherine A. Lippi, Leah R. Johnson, Marco Neira, Jason R. Rohr, Sadie J. Ryan, Van Savage, Marta S. Shocket, Rachel Sippy, Anna M. Stewart Ibarra, Matthew B.

Thomas, and Oswaldo Villena. Thermal biology of mosquito-borne disease. *Ecology Letters*, 22(10):1690–1708, October 2019. ISSN 1461-023X, 1461-0248. doi: 10.1111/ele.13335. URL https://onlinelibrary.wiley.com/doi/10.1111/ele.13335.

[84] M. Trpis. Dry season survival of Aedes aegypti eggs in various breeding sites in the Dar es Salaam area, Tanzania. *Bulletin of the World Health Organization*, 47(3):433–437, 1972. ISSN 0042-9686.

[85] Amy Wesolowski, Taimur Qureshi, Maciej F. Boni, Pål Roe Sundsøy, Michael A. Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O. Buckee. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences*, 112(38):11887–11892, September 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1504964112. URL https://pnas.org/doi/full/10.1073/pnas.1504964112.

[86] Rafael Bomfim, Sen Pei, Jeffrey Shaman, Teresa Yamana, Hernán A. Makse, José S. Andrade, Antonio S. Lima Neto, and Vasco Furtado. Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. *Journal of The Royal Society Interface*, 17(171):20200691, October 2020. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2020.0691. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0691.

[87] Erin A. Mordecai, Krijn P. Paaijmans, Leah R. Johnson, Christian Balzer, Tal Ben-Horin, Emily De Moor, Amy McNally, Samraat Pawar, Sadie J. Ryan, Thomas C. Smith, and Kevin D. Lafferty. Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology Letters*, 16(1):22–30, January 2013. ISSN 1461-023X, 1461-0248. doi: 10.1111/ele.12015. URL https://onlinelibrary.wiley.com/doi/10.1111/ele.12015.

[88] John H. Huber, Marissa L. Childs, Jamie M. Caldwell, and Erin A. Mordecai. Seasonal temperature variation influences climate suitability for dengue, chikungunya, and Zika transmission. *PLOS Neglected Tropical Diseases*, 12 (5):e0006451, May 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006451. URL https://dx.plos.org/10.1371/journal.pntd.0006451.

[89] Yuyang Chen, Yiting Xu, Lin Wang, Yilin Liang, Naizhe Li, José Lourenço, Yun Yang, Qiushi Lin, Ligui Wang, He Zhao, Bernard Cazelles, Hongbin Song, Ziyan Liu, Zengmiao Wang, Oliver J. Brady, Simon Cauchemez, and Huaiyu Tian. Indian Ocean temperature anomalies predict long-term global dengue trends. *Science*, 384(6696): 639–646, May 2024. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adj4427. URL https://www.science.org/doi/10.1126/science.adj4427.

[90] Henrik Salje. Quantifying the effect of vector targeted interventions to control dengue. *The Lancet Infectious Diseases*, 25(2):139–140, February 2025. ISSN 1473-3099, 1474-4457. doi: 10.1016/S1473-3099(24)00533-4. URL https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(24)00533-4/fulltext. Publisher: Elsevier.

[91] Carlos Alberto Montenegro-Quiñonez, Valérie R. Louis, Olaf Horstick, Raman Velayudhan, Peter Dambach, and Silvia Runge-Ranzinger. Interventions against Aedes/dengue at the household level: a systematic review and meta-analysis. *eBioMedicine*, 93, July 2023. ISSN 2352-3964. doi: 10.1016/j.ebiom.2023.104660. URL https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00225-6/fulltext. Publisher: Elsevier.

[92] Ali Jaffal, Johanna Fite, Thierry Baldet, Pascal Delaunay, Frédéric Jourdain, Ronald Mora-Castillo, Marie-Marie Olive, and David Roiz. Current evidences of the efficacy of mosquito mass-trapping interventions to reduce Aedes aegypti and Aedes albopictus populations and Aedes-borne virus transmission. *PLOS Neglected Tropical Diseases*, 17(3):e0011153, March 2023. ISSN 1935-2735. doi: 10.1371/journal.pntd.0011153. URL https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0011153. Publisher: Public Library of Science.

[93] Leigh R. Bowman, Sarah Donegan, and Philip J. McCall. Is Dengue Vector Control Deficient in Effectiveness or Evidence?: Systematic Review and Meta-analysis. *PLOS Neglected Tropical Diseases*, 10(3):e0004551, March 2016. ISSN 1935-2735. doi: 10.1371/journal.pntd.0004551. URL https://dx.plos.org/10.1371/journal.pntd.0004551.

[94] Adi Utarini, Citra Indriani, Riris A. Ahmad, Warsito Tantowijoyo, Eggi Arguni, M. Ridwan Ansari, Endah Supriyati, D. Satria Wardana, Yeti Meitika, Inggrid Ernesia, Indah Nurhayati, Equatori Prabowo, Bekti Andari, Benjamin R. Green, Lauren Hodgson, Zoe Cutcher, Edwige Rancès, Peter A. Ryan, Scott L. O'Neill, Suzanne M. Dufault, Stephanie K. Tanamas, Nicholas P. Jewell, Katherine L. Anders, and Cameron P. Simmons. Efficacy of Wolbachia-Infected Mosquito Deployments for the Control of Dengue. *New England Journal of Medicine*, 384(23):2177–2186, June 2021. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMoa2030243. URL http://www.nejm.org/doi/10.1056/NEJMoa2030243.

[95] Marie-Marie Olive, Gilbert Le Goff, Thierry Baldet, and David Roiz. Are the sterile insect technique and the incompatible insect techniques effective in reducing Aedes mosquito populations?, February 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.02.03.24302193.

[96] David Roiz, Anne L. Wilson, Thomas W. Scott, Dina M. Fonseca, Frédéric Jourdain, Pie Müller, Raman Velayudhan, and Vincent Corbel. Integrated Aedes management for the control of Aedes-borne diseases. *PLOS Neglected Tropical Diseases*, 12(12):e0006845, December 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006845. URL https://dx.plos.org/10.1371/journal.pntd.0006845.

[97] Abhishek Lachyan, Rafdzah Ahmad Zaki, Bratati Banerjee, and Nasrin Aghamohammadi. The Effect of Community-Based Intervention on Dengue Awareness and Prevention Among Poor Urban Communities in Delhi, India. *Journal of Research in Health Sciences*, 23(4):e00596, December 2023. ISSN 2228-7809. doi: 10.34172/jrhs.2023.131.

[98] Ute Deichmann. From Gregor Mendel to Eric Davidson: Mathematical Models and Basic Principles in Biology. *Journal of Computational Biology*, 26(7):637–652, July 2019. ISSN 1066-5277. doi: 10.1089/cmb.2019.0087. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6763957/.

[99] Hans Heesterbeek, Roy M. Anderson, Viggo Andreasen, Shweta Bansal, Daniela De Angelis, Chris Dye, Ken T. D. Eames, W. John Edmunds, Simon D. W. Frost, Sebastian Funk, T. Deirdre Hollingsworth, Thomas House, Valerie Isham, Petra Klepac, Justin Lessler, James O. Lloyd-Smith, C. Jessica E. Metcalf, Denis Mollison, Lorenzo Pellis, Juliet R. C. Pulliam, Mick G. Roberts, Cecile Viboud, and Isaac Newton Institute IDD Collaboration. Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227):aaa4339, March 2015. doi: 10.1126/science.aaa4339. URL https://www.science.org/doi/full/10.1126/science.aaa4339. Publisher: American Association for the Advancement of Science.

[100] Alexander D Becker, Kyra H Grantz, Sonia T Hegde, Sophie Bérubé, Derek A T Cummings, and Amy Wesolowski. Development and dissemination of infectious disease dynamic transmission models during the COVID-19 pandemic: what can we learn from other pathogens and how can we move forward? *The Lancet Digital Health*, 3(1):e41–e50, January 2021. ISSN 25897500. doi: 10.1016/S2589-7500(20)30268-5. URL https://linkinghub.elsevier.com/retrieve/pii/S2589750020302685.

[101] Andreia Vasconcelos, Jonathan D King, Cláudio Nunes-Alves, Roy Anderson, Daniel Argaw, Maria-Gloria Basáñez, Shakir Bilal, David J Blok, Seth Blumberg, Anna Borlase, Oliver J Brady, Raiha Browning, Nakul Chitnis, Luc E Coffeng, Emily H Crowley, Zulma M Cucunubá, Derek A T Cummings, Christopher Neil Davis, Emma Louise Davis, Matthew Dixon, Andrew Dobson, Louise Dyson, Michael French, Claudio Fronterre, Emanuele Giorgi, Ching-I

Huang, Saurabh Jain, Ananthu James, Sung Hye Kim, Klodeta Kura, Ana Lucianez, Michael Marks, Pamela Sabina Mbabazi, Graham F Medley, Edwin Michael, Antonio Montresor, Nyamai Mutono, Thumbi S Mwangi, Kat S Rock, Martha-Idalí Saboyá-Díaz, Misaki Sasanami, Markus Schwehm, Simon E F Spencer, Ariktha Srivathsan, Robert S Stawski, Wilma A Stolk, Samuel A Sutherland, Louis-Albert Tchuem Tchuenté, Sake J De Vlas, Martin Walker, Simon J Brooker, T Déirdre Hollingsworth, Anthony W Solomon, and Ibrahima Socé Fall. Accelerating Progress Towards the 2030 Neglected Tropical Diseases Targets: How Can Quantitative Modeling Support Programmatic Decisions? *Clinical Infectious Diseases*, 78(Supplement_2):S83–S92, April 2024. ISSN 1058-4838, 1537-6591. doi: 10.1093/cid/ciae082. URL https://academic.oup.com/cid/article/78/Supplement_2/S83/7657829.

[102] Eamon Conway, Camelia R. Walker, Christopher Baker, Michael J. Lydeamore, Gerard E. Ryan, Trish Campbell, Joel C. Miller, Nic Rebuli, Max Yeung, Greg Kabashima, Nicholas Geard, James Wood, James M. McCaw, Jodie McVernon, Nick Golding, David J. Price, and Freya M. Shearer. COVID-19 vaccine coverage targets to inform reopening plans in a low incidence setting. *Proceedings of the Royal Society B: Biological Sciences*, 290(2005): 20231437, August 2023. doi: 10.1098/rspb.2023.1437. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2023.1437. Publisher: Royal Society.

[103] Inga Holmdahl and Caroline Buckee. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *New England Journal of Medicine*, 383(4):303–305, July 2020. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMp2016822. URL http://www.nejm.org/doi/10.1056/NEJMp2016822.

[104] David L. Smith, T. Alex Perkins, Robert C. Reiner, Jr., Christopher M. Barker, Tianchan Niu, Luis Fernando Chaves, Alicia M. Ellis, Dylan B. George, Arnaud Le Menach, Juliet R. C. Pulliam, Donal Bisanzio, Caroline Buckee, Christinah Chiyaka, Derek A. T. Cummings, Andres J. Garcia, Michelle L. Gatton, Peter W. Gething, David M. Hartley, Geoffrey Johnston, Eili Y. Klein, Edwin Michael, Alun L. Lloyd, David M. Pigott, William K. Reisen, Nick Ruktanonchai, Brajendra K. Singh, Jeremy Stoller, Andrew J. Tatem, Uriel Kitron, H. Charles J. Godfray, Justin M. Cohen, Simon I. Hay, and Thomas W. Scott. Recasting the theory of mosquito-borne pathogen transmission dynamics and control. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 108(4):185–197, April 2014. ISSN 0035-9203. doi: 10.1093/trstmh/tru026. URL https://doi.org/10.1093/trstmh/tru026.

[105] Beate Jahn, Sarah Friedrich, Joachim Behnke, Joachim Engel, Ursula Garczarek, Ralf Münnich, Markus Pauly, Adalbert Wilhelm, Olaf Wolkenhauer, Markus Zwick, Uwe Siebert, and Tim Friede. On the role of data, statistics and decisions in a pandemic. *AStA Advances in Statistical Analysis*, 106(3):349–382, September 2022. ISSN 1863-818X. doi: 10.1007/s10182-022-00439-7. URL https://doi.org/10.1007/s10182-022-00439-7.

[106] Lee A. Segel. *Modeling dynamic phenomena in molecular and cellular biology*. Cambridge University Press, Cambridge ; New York, 1984. ISBN 978-0-521-25465-6 978-0-521-27477-7.

[107] Katelyn M. Gostic, Lauren McGough, Edward B. Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, Rene Niehus, James A. Hay, Pablo M. De Salazar, Joel Hellewell, Sophie Meakin, James D. Munday, Nikos I. Bosse, Katharine Sherrat, Robin N. Thompson, Laura F. White, Jana S. Huisman, Jérémie Scire, Sebastian Bonhoeffer, Tanja Stadler, Jacco Wallinga, Sebastian Funk, Marc Lipsitch, and Sarah Cobey. Practical considerations for measuring the effective reproductive number, R. *PLOS Computational Biology*, 16(12):e1008409, December 2020. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008409. URL https://dx.plos.org/10.1371/journal.pcbi.1008409.

[108] Chris P. Jewell, Alison C. Hale, Barry S. Rowlingson, Christopher Suter, Jonathan M. Read, and Gareth O. Roberts. Bayesian inference for high-dimensional discrete-time epidemic models: spatial dynamics of the UK COVID-19 outbreak, July 2023. URL http://arxiv.org/abs/2306.07987. arXiv:2306.07987 [physics].

[109] Akira Endo, Edwin Van Leeuwen, and Marc Baguelin. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics*, 29:100363, December 2019. ISSN 17554365. doi: 10.1016/j.epidem.2019.100363. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436519300301.

[110] Christina Pagel and Christian A. Yates. Tackling the pandemic with (biased) data. *Science*, 374(6566):403–404, October 2021. doi: 10.1126/science.abi6602. URL https://www.science.org/doi/10.1126/science.abi6602. Publisher: American Association for the Advancement of Science.

[111] Kit Gallagher, Richard Creswell, David Gavaghan, and Ben Lambert. Identification and attribution of weekly periodic biases in global epidemiological time series data. *BMC Research Notes*, 18(1):78, February 2025. ISSN 1756-0500. doi: 10.1186/s13104-025-07145-y. URL https://doi.org/10.1186/s13104-025-07145-y.

[112] Rinette Badker, Kierste Miller, Chris Pardee, Ben Oppenheim, Nicole Stephenson, Benjamin Ash, Tanya Philippsen, Christopher Ngoon, Partrick Savage, Cathine Lam, and Nita Madhav. Challenges in reported COVID-19 data: best practices and recommendations for future epidemics. *BMJ Global Health*, 6(5), May 2021. ISSN 2059-7908. doi: 10.1136/bmjgh-2021-005542. URL https://gh.bmj.com/content/6/5/e005542. Publisher: BMJ Publishing Group Ltd.

[113] Sang Woo Park, Andrei R. Akhmetzhanov, Kelly Charniga, Anne Cori, Nicholas G. Davies, Jonathan Dushoff, Sebastian Funk, Katie Gostic, Bryan Grenfell, Natalie M. Linton, Marc Lipsitch, Adrian Lison, Christopher E. Overton, Thomas Ward, and Sam Abbott. Estimating epidemiological delay distributions for infectious diseases, January 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.01.12.24301247.

[114] Kelly Charniga, Sang Woo Park, Andrei R Akhmetzhanov, Anne Cori, Jonathan Dushoff, Sebastian Funk, Katelyn M Gostic, Natalie M Linton, Adrian Lison, Christopher E Overton, Juliet R C Pulliam, Thomas Ward, Simon Cauchemez, and Sam Abbott. Best practices for estimating and reporting epidemiological delay distributions of infectious diseases using public health surveillance and healthcare data, 2024. URL https://arxiv.org/abs/2405.08841. Version Number: 1.

[115] Katharine Sherratt, Sam Abbott, Sophie R. Meakin, Joel Hellewell, James D. Munday, Nikos Bosse, null null, Mark Jit, and Sebastian Funk. Exploring surveillance data biases when estimating the reproduction number: with insights into subpopulation transmission of COVID-19 in England. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829):20200283, May 2021. doi: 10.1098/rstb.2020.0283. URL https://royalsocietypublishing.org/doi/full/10.1098/rstb.2020.0283. Publisher: Royal Society.

[116] Odo Diekmann and J. A. P. Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Wiley series in mathematical and computational biology. Wiley, Chichester Weinheim, 2000. ISBN 978-0-471-98682-9 978-0-471-49241-2.

[117] Roy M. Anderson and Robert M. May. *Infectious diseases of humans: dynamics and control*. Oxford Univ. Press, Oxford, reprinted edition, 2010. ISBN 978-0-19-854040-3.

[118] Carolin Vegvari, Sam Abbott, Frank Ball, Ellen Brooks-Pollock, Robert Challen, Benjamin S Collyer, Ciara Dangerfield, Julia R Gog, Katelyn M Gostic, Jane M Heffernan, T Déirdre Hollingsworth, Valerie Isham, Eben Kenah, Denis Mollison, Jasmina Panovska-Griffiths, Lorenzo Pellis, Michael G Roberts, Gianpaolo Scalia Tomba, Robin N Thompson, and Pieter Trapman. Commentary on the use of the reproduction number *R* during the COVID-19 pandemic. *Statistical Methods in Medical Research*, 31(9):1675–1685, September 2022. ISSN 0962-2802, 1477-0334. doi: 10.1177/09622802211037079. URL http://journals.sagepub.com/doi/10.1177/09622802211037079.

[119] Andrew F. Brouwer. Why the Spectral Radius? An intuition-building introduction to the basic reproduction number. *Bulletin of Mathematical Biology*, 84(9):96, September 2022. ISSN 0092-8240, 1522-9602. doi: 10.1007/s11538-022-01057-9. URL https://link.springer.com/10.1007/s11538-022-01057-9.

[120] Nienke Hartemink, Daniela Cianci, and Paul Reiter. R0 for Vector-Borne Diseases: Impact of the Assumption for the Duration of the Extrinsic Incubation Period. *Vector Borne and Zoonotic Diseases*, 15(3):215–217, March 2015. ISSN 1530-3667. doi: 10.1089/vbz.2014.1684. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4369930/.

[121] David L. Smith, Katherine E. Battle, Simon I. Hay, Christopher M. Barker, Thomas W. Scott, and F. Ellis McKenzie. Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens. *PLOS Pathogens*, 8(4):e1002588, April 2012. ISSN 1553-7374. doi: 10.1371/journal.ppat.1002588. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002588. Publisher: Public Library of Science.

[122] Matt J. Keeling, Matthew James Keeling, and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton Univ. Press, Princeton, NJ, 2008. ISBN 978-0-691-11617-4.

[123] Ottar N. Bjørnstad. *Epidemics: Models and Data using R*. Use R! Springer International Publishing, Cham, 2018. ISBN 978-3-319-97486-6 978-3-319-97487-3. doi: 10.1007/978-3-319-97487-3. URL http://link.springer.com/10.1007/978-3-319-97487-3.

[124] Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2):taaa021, March 2020. ISSN 1708-8305. doi: 10.1093/jtm/taaa021. URL https://academic.oup.com/jtm/article/doi/10.1093/jtm/taaa021/5735319.

[125] Kevin Linka, Mathias Peirlinck, and Ellen Kuhl. The reproduction number of COVID-19 and its correlation with public health interventions. *medRxiv*, page 2020.05.01.20088047, July 2020. doi: 10.1101/2020.05.01.20088047. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7359536/.

[126] Ying Liu and Joacim Rocklöv. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *Journal of Travel Medicine*, 28(7):taab124, October 2021. ISSN 1708-8305. doi: 10.1093/jtm/taab124. URL https://doi.org/10.1093/jtm/taab124.

[127] Ying Liu, Kate Lillepold, Jan C. Semenza, Yesim Tozan, Mikkel B. M. Quam, and Joacim Rocklöv. Reviewing estimates of the basic reproduction number for dengue, Zika and chikungunya across global climate zones. *Environmental Research*, 182:109114, March 2020. ISSN 0013-9351. doi: 10.1016/j.envres.2020.109114. URL https://www.sciencedirect.com/science/article/pii/S0013935120300050.

[128] O. Diekmann, J. A. Heesterbeek, and J. A. Metz. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990. ISSN 0303-6812. doi: 10.1007/BF00178324.

[129] Kris V. Parag. How to Measure the Controllability of an Infectious Disease? *Physical Review X*, 14(3):031041, September 2024. ISSN 2160-3308. doi: 10.1103/PhysRevX.14.031041. URL https://link.aps.org/doi/10.1103/PhysRevX.14.031041.

[130] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, November 2005. ISSN 1476-4687. doi: 10.1038/nature04153. URL https://www.nature.com/articles/nature04153. Publisher: Nature Publishing Group.

[131] T Déirdre Hollingsworth. Controlling infectious disease outbreaks: Lessons from mathematical modelling. *Journal of Public Health Policy*, 30(3):328–341, 2009. ISSN 0197-5897. doi: 10.1057/jphp.2009.13. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7099230/.

[132] Nicholas P. Jewell and Joseph A. Lewnard. On the use of the Reproduction Number for SARS-COV-2: Estimation, Misinterpretations and Relationships with other Ecological Measures. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(Supplement_1):S16–S27, November 2022. ISSN 0964-1998, 1467-985X. doi: 10.1111/rssa.12860. URL https://academic.oup.com/jrsssa/article/185/Supplement_1/S16/7069477.

[133] J Wallinga and M Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, February 2007. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2006.3754. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2006.3754.

[134] Kris V. Parag, Robin N. Thompson, and Christl A. Donnelly. Are epidemic growth rates more informative than reproduction numbers? *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185:S5–S15, November 2022. ISSN 0964-1998, 1467-985X. doi: 10.1111/rssa.12867. URL https://academic.oup.com/jrsssa/article/185/Supplement_1/S5/7069470.

[135] Jonathan Dushoff and Sang Woo Park. Speed and strength of an epidemic intervention. *Proceedings of the Royal Society B: Biological Sciences*, 288(1947):rspb.2020.1556, 20201556, March 2021. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2020.1556. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2020.1556.

[136] Roy M. Anderson, Christl A. Donnelly, T. Deirdre Hollingsworth, Matt J. Keeling, Carolin Vegvari, Rebecca Baggaley, and Rosie Maddren. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. Technical report, The Royal Society, August 2020. URL https://royalsociety.org/-/media/policy/projects/set-c/set-covid-19-r-estimates.pdf.

[137] Christophe Fraser. Estimating individual and household reproduction numbers in an emerging epidemic. *PLOS ONE*, 2(8):e758, August 2007. ISSN 1932-6203. doi: 10.1371/journal.pone.0000758. URL https://dx.plos.org/10.1371/journal.pone.0000758.

[138] J. Wallinga. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *American Journal of Epidemiology*, 160(6):509–516, September 2004. ISSN 0002-9262. doi: 10.1093/aje/kwh255. URL https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwh255.

[139] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, November 2013. ISSN 1476-6256, 0002-9262. doi: 10.1093/aje/kwt133. URL https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwt133.

[140] R.N. Thompson, J.E. Stockwin, R.D. Van Gaalen, J.A. Polonsky, Z.N. Kamvar, P.A. Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart, J. Lessler, S. Cauchemez, and A. Cori. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, December 2019. ISSN 17554365. doi: 10.1016/j.epidem.2019.100356. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436519300350.

[141] Nicholas Steyn and Kris V. Parag. Robust uncertainty quantification in popular estimators of the instantaneous reproduction number, January 2025. URL https://www.medrxiv.org/content/10.1101/2024.10.22.24315918v2. Pages: 2024.10.22.24315918.

[142] Kris V. Parag, Benjamin J. Cowling, and Ben Lambert. Angular reproduction numbers improve estimates of transmissibility when disease generation times are misspecified or time-varying. *Proceedings of the Royal Society B: Biological Sciences*, 290(2007):20231664, September 2023. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2023.1664. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2023.1664.

[143] A.L. Lloyd. Sensitivity of Model-Based Epidemiological Parameter Estimation to Model Assumptions. *Mathematical and Statistical Estimation Approaches in Epidemiology*, pages 123–141, 2009. doi: 10.1007/978-90-481-2313-1_6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7121564/.

[144] Sumali Bajaj, Robin Thompson, and Ben Lambert. A renewal-equation approach to estimating Rt and infectious disease case counts in the presence of reporting delays. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 383(2292):20240357, March 2025. doi: 10.1098/rsta.2024.0357. URL https://royalsocietypublishing.org/doi/full/10.1098/rsta.2024.0357. Publisher: Royal Society.

[145] Sang Woo Park, Benjamin M. Bolker, Sebastian Funk, C. Jessica E. Metcalf, Joshua S. Weitz, Bryan T. Grenfell, and Jonathan Dushoff. The importance of the generation interval in investigating dynamics and control of new SARS-CoV-2 variants. *Journal of The Royal Society Interface*, 19(191):20220173, June 2022. ISSN 1742-5662. doi: 10.1098/rsif.2022.0173. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2022.0173.

[146] William S Hart, Sam Abbott, Akira Endo, Joel Hellewell, Elizabeth Miller, Nick Andrews, Philip K Maini, Sebastian Funk, and Robin N Thompson. Inference of the SARS-CoV-2 generation time using UK household data. *eLife*, 11:e70767, February 2022. ISSN 2050-084X. doi: 10.7554/eLife.70767. URL https://elifesciences.org/articles/70767.

[147] David Champredon and Jonathan Dushoff. Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821):20152026, December 2015. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2015.2026. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2015.2026.

[148] Sam Abbott, Joel Hellewell, Robin N. Thompson, Katharine Sherratt, Hamish P. Gibbs, Nikos I. Bosse, James D. Munday, Sophie Meakin, Emma L. Doughty, June Young Chun, Yung-Wai Desmond Chan, Flavio Finger, Paul Campbell, Akira Endo, Carl A. B. Pearson, Amy Gimma, Tim Russell, CMMID COVID modelling group, Stefan Flasche, Adam J. Kucharski, Rosalind M. Eggo, and Sebastian Funk. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*, 5:112, December 2020. ISSN 2398-502X. doi: 10.12688/wellcomeopenres.16006.2. URL https://wellcomeopenresearch.org/articles/5-112/v2.

[149] Kris V. Parag. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLOS Computational Biology*, 17(9):e1009347, September 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009347. URL https://dx.plos.org/10.1371/journal.pcbi.1009347.

[150] B. Boldin, O. Diekmann, and J. A. J. Metz. Population growth in discrete time: a renewal equation oriented survey. *Journal of Difference Equations and Applications*, pages 1–29, October 2023. ISSN 1023-6198, 1563-5120. doi: 10.1080/10236198.2023.2265499. URL https://www.tandfonline.com/doi/full/10.1080/10236198.2023.2265499.

[151] Eugenia Franco, Odo Diekmann, and Mats Gyllenberg. Modelling physiologically structured populations: renewal equations and partial differential equations. *Journal of Evolution Equations*, 23(3):46, September 2023. ISSN 1424-3199, 1424-3202. doi: 10.1007/s00028-023-00880-4. URL https://link.springer.com/10.1007/s00028-023-00880-4.

[152] Mikko S. Pakkanen, Xenia Miscouridou, Matthew J. Penn, Charles Whittaker, Tresnia Berah, Swapnil Mishra, Thomas A. Mellan, and Samir Bhatt. Unifying incidence and prevalence under a time-varying general branching process. *Journal of Mathematical Biology*, 87(2):35, August 2023. ISSN 0303-6812, 1432-1416. doi: 10.1007/s00285-023-01958-w. URL https://link.springer.com/10.1007/s00285-023-01958-w.

[153] Lorenzo Pellis, Simon Cauchemez, Neil M. Ferguson, and Christophe Fraser. Systematic selection between age and household structure for models aimed at emerging epidemic predictions. *Nature Communications*, 11(1):906, February 2020. ISSN 2041-1723. doi: 10.1038/s41467-019-14229-4. URL https://www.nature.com/articles/s41467-019-14229-4.

[154] Nina L. Dennington, Marissa K. Grossman, Janet L. Teeple, Leah R. Johnson, Marta S. Shocket, Elizabeth A. McGraw, and Matthew B. Thomas. Local adaptation of the mosquito vector, Aedes aegypti, and implications for predicting the effects of temperature and climate change on dengue transmission, February 2025. URL https://www.biorxiv.org/content/10.1101/2025.02.14.637444v1. Pages: 2025.02.14.637444 Section: New Results.

[155] Souvik Chakraborty, Emily Zigmond, Sher Shah, Diya Dayal, Massamba Sylla, Jewelna Akorli, Sampson Otoo, Noah H. Rose, Carolyn S. McBride, Peter A. Armbruster, and Joshua B. Benoit. Thermal tolerance of *Aedes aegypti* mosquito eggs is associated with urban adaptation and human interactions. *Journal of Thermal Biology*, page 104167, June 2025. ISSN 0306-4565. doi: 10.1016/j.jtherbio.2025.104167. URL https://www.sciencedirect.com/science/article/pii/S030645652500124X.

[156] Amir S. Siraj, Rachel J. Oidtman, John H. Huber, Moritz U. G. Kraemer, Oliver J. Brady, Michael A. Johansson, and T. Alex Perkins. Temperature modulates dengue virus epidemic growth rates through its effects on reproduction numbers and generation intervals. *PLOS Neglected Tropical Diseases*, 11(7):e0005797, July 2017. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005797. URL https://dx.plos.org/10.1371/journal.pntd.0005797.

[157] Leonardo S Bastos, Theodoros Economou, Marcelo F C Gomes, Daniel A M Villela, Flavio C Coelho, Oswaldo G Cruz, Oliver Stoner, Trevor Bailey, and Claudia T Codeço. A modelling approach for correcting reporting delays in disease surveillance data. *Statistics in Medicine*, 38(22):4363–4377, September 2019. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.8303. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.8303.

[158] Michael C. Runge, Katriona Shea, Emily Howerton, Katie Yan, Harry Hochheiser, Erik Rosenstrom, William J.M. Probert, Rebecca Borchering, Madhav V. Marathe, Bryan Lewis, Srinivasan Venkatramanan, Shaun Truelove, Justin Lessler, and Cécile Viboud. Scenario design for infectious disease projections: Integrating concepts from decision analysis and experimental design. *Epidemics*, 47:100775, June 2024. ISSN 17554365. doi: 10.1016/j.epidem.2024.100775. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436524000367.

[159] Emilie Finch, Martín Lotto Batista, Tilly Alcayna, Sophie A. Lee, Isabel K. Fletcher, and Rachel Lowe. Chapter 12 Early warning systems for vector-borne diseases: engagement, methods and implementation. pages 347 – 386. Wageningen Academic, Leiden, The Netherlands, 2023. ISBN 978-90-04-68865-0. doi: 10.3920/9789004688650_014. URL https://brill.com/view/book/9789004688650/BP000023.xml.

[160] Loïc Berger, Nicolas Berger, Valentina Bosetti, Itzhak Gilboa, Lars Peter Hansen, Christopher Jarvis, Massimo Marinacci, and Richard D. Smith. Rational policymaking during a pandemic. *Proceedings of the National Academy of Sciences*, 118(4):e2012704118, January 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2012704118. URL https://pnas.org/doi/full/10.1073/pnas.2012704118.

[161] Allan H. Murphy. What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8(2):281–293, June 1993. ISSN 0882-8156, 1520-0434. doi: 10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2. URL http://journals.ametsoc.org/doi/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

[162] Felipe J. Colón-González, Leonardo Soares Bastos, Barbara Hofmann, Alison Hopkin, Quillon Harpham, Tom Crocker, Rosanna Amato, Iacopo Ferrario, Francesca Moschini, Samuel James, Sajni Malde, Eleanor Ainscoe, Vu Sinh Nam, Dang Quang Tan, Nguyen Duc Khoa, Mark Harrison, Gina Tsarouchi, Darren Lumbroso, Oliver J. Brady, and Rachel Lowe. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Medicine*, 18(3):e1003542, March 2021. ISSN 1549-1676. doi: 10.1371/journal.pmed.1003542. URL https://dx.plos.org/10.1371/journal.pmed.1003542.

[163] Angel N. Desai, Moritz U. G. Kraemer, Sangeeta Bhatia, Anne Cori, Pierre Nouvellet, Mark Herringer, Emily L. Cohn, Malwina Carrion, John S. Brownstein, Lawrence C. Madoff, and Britta Lassmann. Real-time Epidemic Forecasting: Challenges and Opportunities. *Health Security*, 17(4):268–275, August 2019. ISSN 2326-5094. doi: 10.1089/hs.2019.0022. URL https://www.liebertpub.com/doi/10.1089/hs.2019.0022. Publisher: Mary Ann Liebert, Inc., publishers.

[164] Steve Brooks, editor. *Handbook for Markov chain Monte Carlo*. Taylor & Francis, Boca Raton, 2011. ISBN 978-1-4200-7941-8.

[165] Christopher J. M. Whitty. What makes an academic paper useful for health policy? *BMC Medicine*, 13(1):301, December 2015. ISSN 1741-7015. doi: 10.1186/s12916-015-0544-8. URL https://doi.org/10.1186/s12916-015-0544-8.

[166] Samuel V. Scarpino and Giovanni Petri. On the predictability of infectious disease outbreaks. *Nature Communications*, 10(1):898, February 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-08616-0. URL https://www.nature.com/articles/s41467-019-08616-0.

[167] Sarabeth M. Mathis, Alexander E. Webber, Tomás M. León, Erin L. Murray, Monica Sun, Lauren A. White, Logan C. Brooks, Alden Green, Addison J. Hu, Roni Rosenfeld, Dmitry Shemetov, Ryan J. Tibshirani, Daniel J. McDonald, Sasikiran Kandula, Sen Pei, Rami Yaari, Teresa K. Yamana, Jeffrey Shaman, Pulak Agarwal, Srikar Balusu, Gautham Gururajan, Harshavardhan Kamarthi, B. Aditya Prakash, Rishi Raman, Zhiyuan Zhao, Alexander Rodríguez, Akilan Meiyappan, Shalina Omar, Prasith Baccam, Heidi L. Gurung, Brad T. Suchoski, Steve A. Stage, Marco Ajelli, Allisandra G. Kummer, Maria Litvinova, Paulo C. Ventura, Spencer Wadsworth, Jarad Niemi, Erica Carcelen, Alison L. Hill, Sara L. Loo, Clifton D. McKee, Koji Sato, Claire Smith, Shaun Truelove, Sung-mok Jung, Joseph C. Lemaitre, Justin Lessler, Thomas McAndrew, Wenxuan Ye, Nikos Bosse, William S. Hlavacek, Yen Ting Lin, Abhishek Mallela, Graham C. Gibson, Ye Chen, Shelby M. Lamm, Jaechoul Lee, Richard G. Posner, Amanda C. Perofsky, Cécile Viboud, Leonardo Clemente, Fred Lu, Austin G. Meyer, Mauricio Santillana, Matteo Chinazzi, Jessica T. Davis, Kunpeng Mu, Ana Pastore Y Piontti, Alessandro Vespignani, Xinyue Xiong, Michal Ben-Nun, Pete Riley, James Turtle, Chis Hulme-Lowe, Shakeel Jessa, V. P. Nagraj, Stephen D. Turner, Desiree Williams, Avranil Basu, John M. Drake, Spencer J. Fox, Ehsan Suez, Monica G. Cojocaru, Edward W. Thommes, Estee Y. Cramer, Aaron

Gerding, Ariane Stark, Evan L. Ray, Nicholas G. Reich, Li Shandross, Nutcha Wattanachit, Yijin Wang, Martha W. Zorn, Majd Al Aawar, Ajitesh Srivastava, Lauren A. Meyers, Aniruddha Adiga, Benjamin Hurt, Gursharn Kaur, Bryan L. Lewis, Madhav Marathe, Srinivasan Venkatramanan, Patrick Butler, Andrew Farabow, Naren Ramakrishnan, Nikhil Muralidhar, Carrie Reed, Matthew Biggerstaff, and Rebecca K. Borchering. Evaluation of FluSight influenza forecasting in the 2021–22 and 2022–23 seasons with a new target laboratory-confirmed influenza hospitalizations. *Nature Communications*, 15(1):6289, July 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-50601-9. URL https://www.nature.com/articles/s41467-024-50601-9.

[168] Rachel Lowe, Trevor C. Bailey, David B. Stephenson, Tim E. Jupp, Richard J. Graham, Christovam Barcellos, and Marilia Sá Carvalho. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in Medicine*, 32(5):864–883, February 2013. ISSN 02776715. doi: 10.1002/sim.5549. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.5549.

[169] Michael A. Johansson, Karyn M. Apfeldorf, Scott Dobson, Jason Devita, Anna L. Buczak, Benjamin Baugher, Linda J. Moniz, Thomas Bagley, Steven M. Babin, Erhan Guven, Teresa K. Yamana, Jeffrey Shaman, Terry Moschou, Nick Lothian, Aaron Lane, Grant Osborne, Gao Jiang, Logan C. Brooks, David C. Farrow, Sangwon Hyun, Ryan J. Tibshirani, Roni Rosenfeld, Justin Lessler, Nicholas G. Reich, Derek A. T. Cummings, Stephen A. Lauer, Sean M. Moore, Hannah E. Clapham, Rachel Lowe, Trevor C. Bailey, Markel García-Díez, Marilia Sá Carvalho, Xavier Rodó, Tridip Sardar, Richard Paul, Evan L. Ray, Krzysztof Sakrejda, Alexandria C. Brown, Xi Meng, Osonde Osoba, Raffaele Vardavas, David Manheim, Melinda Moore, Dhananjai M. Rao, Travis C. Porco, Sarah Ackley, Fengchen Liu, Lee Worden, Matteo Convertino, Yang Liu, Abraham Reddy, Eloy Ortiz, Jorge Rivero, Humberto Brito, Alicia Juarrero, Leah R. Johnson, Robert B. Gramacy, Jeremy M. Cohen, Erin A. Mordecai, Courtney C. Murdock, Jason R. Rohr, Sadie J. Ryan, Anna M. Stewart-Ibarra, Daniel P. Weikel, Antarpreet Jutla, Rakibul Khan, Marissa Poultney, Rita R. Colwell, Brenda Rivera-García, Christopher M. Barker, Jesse E. Bell, Matthew Biggerstaff, David Swerdlow, Luis Mier-y Teran-Romero, Brett M. Forshey, Juli Trtanj, Jason Asher, Matt Clay, Harold S. Margolis, Andrew M. Hebbeler, Dylan George, and Jean-Paul Chretien. An open challenge to advance probabilistic forecasting for dengue epidemics. *Proceedings of the National Academy of Sciences*, 116(48):24268–24274, November 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1909865116. URL https://pnas.org/doi/full/10.1073/pnas.1909865116.

[170] Velma K. Lopez, Estee Y. Cramer, Robert Pagano, John M. Drake, Eamon B. O'Dea, Madeline Adee, Turgay Ayer, Jagpreet Chhatwal, Ozden O. Dalgic, Mary A. Ladd, Benjamin P. Linas, Peter P. Mueller, Jade Xiao, Johannes Bracher, Alvaro J. Castro Rivadeneira, Aaron Gerding, Tilmann Gneiting, Yuxin Huang, Dasuni Jayawardena, Abdul H. Kanji, Khoa Le, Anja Mühlemann, Jarad Niemi, Evan L. Ray, Ariane Stark, Yijin Wang, Nutcha Wattanachit, Martha W. Zorn, Sen Pei, Jeffrey Shaman, Teresa K. Yamana, Samuel R. Tarasewicz, Daniel J. Wilson, Sid Baccam, Heidi Gurung, Steve Stage, Brad Suchoski, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Guannan Wang, Lily Wang, Yueying Wang, Shan Yu, Lauren Gardner, Sonia Jindal, Maximilian Marshall, Kristen Nixon, Juan Dent, Alison L. Hill, Joshua Kaminsky, Elizabeth C. Lee, Joseph C. Lemaitre, Justin Lessler, Claire P. Smith, Shaun Truelove, Matt Kinsey, Luke C. Mullany, Kaitlin Rainwater-Lovett, Lauren Shin, Katharine Tallaksen, Shelby Wilson, Dean Karlen, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dave Osthus, Jiang Bian, Wei Cao, Zhifeng Gao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Matteo Chinazzi, Jessica T. Davis, Kunpeng Mu, Ana Pastore Y Piontti, Alessandro Vespignani, Xinyue Xiong, Robert Walraven, Jinghui Chen, Quanquan Gu, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Graham Casey Gibson, Daniel Sheldon, Ajitesh Srivastava, Aniruddha Adiga, Benjamin Hurt, Gursharn Kaur, Bryan Lewis, Madhav Marathe, Akhil Sai Peddireddy, Przemyslaw Porebski, Srinivasan Venkatramanan, Lijing Wang, Pragati V. Prasad, Jo W. Walker, Alexander E. Webber, Rachel B. Slayton, Matthew Biggerstaff, Nicholas G. Reich, and Michael A. Johansson. Challenges of COVID-19 Case

Forecasting in the US, 2020–2021. *PLOS Computational Biology*, 20(5):e1011200, May 2024. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011200. URL https://dx.plos.org/10.1371/journal.pcbi.1011200.

[171] Cécile Viboud, Kaiyuan Sun, Robert Gaffey, Marco Ajelli, Laura Fumanelli, Stefano Merler, Qian Zhang, Gerardo Chowell, Lone Simonsen, and Alessandro Vespignani. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21, March 2018. ISSN 17554365. doi: 10.1016/j.epidem.2017.08.002. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436517301275.

[172] Estee Y. Cramer, Evan L. Ray, Velma K. Lopez, Johannes Bracher, Andrea Brennen, Alvaro J. Castro Rivadeneira, Aaron Gerding, Tilmann Gneiting, Katie H. House, Yuxin Huang, Dasuni Jayawardena, Abdul H. Kanji, Ayush Khandelwal, Khoa Le, Anja Mühlemann, Jarad Niemi, Apurv Shah, Ariane Stark, Yijin Wang, Nutcha Wattanachit, Martha W. Zorn, Youyang Gu, Sansiddh Jain, Nayana Bannur, Ayush Deva, Mihir Kulkarni, Srujana Merugu, Alpan Raval, Siddhant Shingi, Avtansh Tiwari, Jerome White, Neil F. Abernethy, Spencer Woody, Maytal Dahan, Spencer Fox, Kelly Gaither, Michael Lachmann, Lauren Ancel Meyers, James G. Scott, Mauricio Tec, Ajitesh Srivastava, Glover E. George, Jeffrey C. Cegan, Ian D. Dettwiller, William P. England, Matthew W. Farthing, Robert H. Hunter, Brandon Lafferty, Igor Linkov, Michael L. Mayo, Matthew D. Parno, Michael A. Rowland, Benjamin D. Trump, Yanli Zhang-James, Samuel Chen, Stephen V. Faraone, Jonathan Hess, Christopher P. Morley, Asif Salekin, Dongliang Wang, Sabrina M. Corsetti, Thomas M. Baer, Marisa C. Eisenberg, Karl Falb, Yitao Huang, Emily T. Martin, Ella McCauley, Robert L. Myers, Tom Schwarz, Daniel Sheldon, Graham Casey Gibson, Rose Yu, Liyao Gao, Yian Ma, Dongxia Wu, Xifeng Yan, Xiaoyong Jin, Yu-Xiang Wang, YangQuan Chen, Lihong Guo, Yanting Zhao, Quanquan Gu, Jinghui Chen, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Hannah Biegel, Joceline Lega, Steve McConnell, V. P. Nagraj, Stephanie L. Guertin, Christopher Hulme-Lowe, Stephen D. Turner, Yunfeng Shi, Xuegang Ban, Robert Walraven, Qi-Jun Hong, Stanley Kong, Axel van de Walle, James A. Turtle, Michal Ben-Nun, Steven Riley, Pete Riley, Ugur Koyluoglu, David DesRoches, Pedro Forli, Bruce Hamory, Christina Kyriakides, Helen Leis, John Milliken, Michael Moloney, James Morgan, Ninad Nirgudkar, Gokce Ozcan, Noah Piwonka, Matt Ravi, Chris Schrader, Elizabeth Shakhnovich, Daniel Siegel, Ryan Spatz, Chris Stiefeling, Barrie Wilkinson, Alexander Wong, Sean Cavany, Guido España, Sean Moore, Rachel Oidtman, Alex Perkins, David Kraus, Andrea Kraus, Zhifeng Gao, Jiang Bian, Wei Cao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Alessandro Vespignani, Matteo Chinazzi, Jessica T. Davis, Kunpeng Mu, Ana Pastore Y Piontti, Xinyue Xiong, Andrew Zheng, Jackie Baek, Vivek Farias, Andreea Georgescu, Retsef Levi, Deeksha Sinha, Joshua Wilde, Georgia Perakis, Mohammed Amine Bennouna, David Nze-Ndong, Divya Singhvi, Ioannis Spantidakis, Leann Thayaparan, Asterios Tsiourvas, Arnab Sarker, Ali Jadbabaie, Devavrat Shah, Nicolas Della Penna, Leo A. Celi, Saketh Sundar, Russ Wolfinger, Dave Osthus, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dean Karlen, Matt Kinsey, Luke C. Mullany, Kaitlin Rainwater-Lovett, Lauren Shin, Katharine Tallaksen, Shelby Wilson, Elizabeth C. Lee, Juan Dent, Kyra H. Grantz, Alison L. Hill, Joshua Kaminsky, Kathryn Kaminsky, Lindsay T. Keegan, Stephen A. Lauer, Joseph C. Lemaitre, Justin Lessler, Hannah R. Meredith, Javier Perez-Saez, Sam Shah, Claire P. Smith, Shaun A. Truelove, Josh Wills, Maximilian Marshall, Lauren Gardner, Kristen Nixon, John C. Burant, Lily Wang, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Guannan Wang, Yueying Wang, Shan Yu, Robert C. Reiner, Ryan Barber, Emmanuela Gakidou, Simon I. Hay, Steve Lim, Chris Murray, David Pigott, Heidi L. Gurung, Prasith Baccam, Steven A. Stage, Bradley T. Suchoski, B. Aditya Prakash, Bijaya Adhikari, Jiaming Cui, Alexander Rodríguez, Anika Tabassum, Jiajia Xie, Pinar Keskinocak, John Asplund, Arden Baxter, Buse Eylul Oruc, Nicoleta Serban, Sercan O. Arik, Mike Dusenberry, Arkady Epshteyn, Elli Kanal, Long T. Le, Chun-Liang Li, Tomas Pfister, Dario Sava, Rajarishi Sinha, Thomas Tsai, Nate Yoder, Jinsung Yoon, Leyou Zhang, Sam Abbott, Nikos I. Bosse, Sebastian Funk, Joel Hellewell, Sophie R. Meakin, Katharine Sherratt, Mingyuan Zhou, Rahi Kalantari, Teresa K. Yamana, Sen Pei, Jeffrey Shaman, Michael L. Li, Dimitris Bertsimas, Omar Skali Lami, Saksham Soni, Hamza Tazi Bouardi, Turgay Ayer, Madeline Adee, Jagpreet Chhatwal, Ozden O. Dalgic,

Mary A. Ladd, Benjamin P. Linas, Peter Mueller, Jade Xiao, Yuanjia Wang, Qinxia Wang, Shanghong Xie, Donglin Zeng, Alden Green, Jacob Bien, Logan Brooks, Addison J. Hu, Maria Jahja, Daniel McDonald, Balasubramanian Narasimhan, Collin Politsch, Samyak Rajanala, Aaron Rumack, Noah Simon, Ryan J. Tibshirani, Rob Tibshirani, Valerie Ventura, Larry Wasserman, Eamon B. O'Dea, John M. Drake, Robert Pagano, Quoc T. Tran, Lam Si Tung Ho, Huong Huynh, Jo W. Walker, Rachel B. Slayton, Michael A. Johansson, Matthew Biggerstaff, and Nicholas G. Reich. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 119(15):e2113561119, April 2022. ISSN 1091-6490. doi: 10.1073/pnas.2113561119.

[173] Estee Y. Cramer, Yuxin Huang, Yijin Wang, Evan L. Ray, Matthew Cornell, Johannes Bracher, Andrea Brennen, Alvaro J. Castro Rivadeneira, Aaron Gerding, Katie House, Dasuni Jayawardena, Abdul Hannan Kanji, Ayush Khandelwal, Khoa Le, Vidhi Mody, Vrushti Mody, Jarad Niemi, Ariane Stark, Apurv Shah, Nutcha Wattanchit, Martha W. Zorn, Nicholas G. Reich, US COVID-19 Forecast Hub Consortium, Tilmann Gneiting, Anja Mühlemann, Youyang Gu, Yixian Chen, Krishna Chintanippu, Viresh Jivane, Ankita Khurana, Ajay Kumar, Anshul Lakhani, Prakhar Mehrotra, Sujitha Pasumarty, Monika Shrivastav, Jialu You, Nayana Bannur, Ayush Deva, Sansiddh Jain, Mihir Kulkarni, Srujana Merugu, Alpan Raval, Siddhant Shingi, Avtansh Tiwari, Jerome White, Aniruddha Adiga, Benjamin Hurt, Bryan Lewis, Madhav Marathe, Akhil Sai Peddireddy, Przemyslaw Porebski, Srinivasan Venkatramanan, Lijing Wang, Maytal Dahan, Spencer Fox, Kelly Gaither, Michael Lachmann, Lauren Ancel Meyers, James G. Scott, Mauricio Tec, Spencer Woody, Ajitesh Srivastava, Tianjian Xu, Jeffrey C. Cegan, Ian D. Dettwiller, William P. England, Matthew W. Farthing, Glover E. George, Robert H. Hunter, Brandon Lafferty, Igor Linkov, Michael L. Mayo, Matthew D. Parno, Michael A. Rowland, Benjamin D. Trump, Samuel Chen, Stephen V. Faraone, Jonathan Hess, Christopher P. Morley, Asif Salekin, Dongliang Wang, Yanli Zhang-James, Thomas M. Baer, Sabrina M. Corsetti, Marisa C. Eisenberg, Karl Falb, Yitao Huang, Emily T. Martin, Ella McCauley, Robert L. Myers, Tom Schwarz, Graham Casey Gibson, Daniel Sheldon, Liyao Gao, Yian Ma, Dongxia Wu, Rose Yu, Xiaoyong Jin, Yu-Xiang Wang, Xifeng Yan, YangQuan Chen, Lihong Guo, Yanting Zhao, Jinghui Chen, Quanquan Gu, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Ishanu Chattopadhyay, Yi Huang, Guoqing Lu, Ruth Pfeiffer, Timothy Sumner, Dongdong Wang, Liqiang Wang, Shunpu Zhang, Zihang Zou, Hannah Biegel, Joceline Lega, Fazle Hussain, Zeina Khan, Frank Van Bussel, Steve McConnell, Stephanie L Guertin, Christopher Hulme-Lowe, V. P. Nagraj, Stephen D. Turner, Benjamín Bejar, Christine Choirat, Antoine Flahault, Ekaterina Krymova, Gavin Lee, Elisa Manetti, Kristen Namigai, Guillaume Obozinski, Tao Sun, Dorina Thanou, Xuegang Ban, Yunfeng Shi, Robert Walraven, Qi-Jun Hong, Axel Van De Walle, Michal Ben-Nun, Steven Riley, Pete Riley, James Turtle, Duy Cao, Joseph Galasso, Jae H. Cho, Areum Jo, David DesRoches, Pedro Forli, Bruce Hamory, Ugur Koyluoglu, Christina Kyriakides, Helen Leis, John Milliken, Michael Moloney, James Morgan, Ninad Nirgudkar, Gokce Ozcan, Noah Piwonka, Matt Ravi, Chris Schrader, Elizabeth Shakhnovich, Daniel Siegel, Ryan Spatz, Chris Stiefeling, Barrie Wilkinson, Alexander Wong, Sean Cavany, Guido España, Sean Moore, Rachel Oidtman, Alex Perkins, Julie S. Ivy, Maria E. Mayorga, Jessica Mele, Erik T. Rosenstrom, Julie L. Swann, Andrea Kraus, David Kraus, Jiang Bian, Wei Cao, Zhifeng Gao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Matteo Chinazzi, Alessandro Vespignani, Xinyue Xiong, Jessica T. Davis, Kunpeng Mu, Ana Pastore Y Piontti, Jackie Baek, Vivek Farias, Andreea Georgescu, Retsef Levi, Deeksha Sinha, Joshua Wilde, Andrew Zheng, Omar Skali Lami, Amine Bennouna, David Nze Ndong, Georgia Perakis, Divya Singhvi, Ioannis Spantidakis, Leann Thayaparan, Asterios Tsiourvas, Shane Weisberg, Ali Jadbabaie, Arnab Sarker, Devavrat Shah, Leo A. Celi, Nicolas D. Penna, Saketh Sundar, Abraham Berlin, Parth D. Gandhi, Thomas McAndrew, Matthew Piriya, Ye Chen, William Hlavacek, Yen Ting Lin, Abhishek Mallela, Ely Miller, Jacob Neumann, Richard Posner, Russ Wolfinger, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dave Osthus, Daniel Wolffram, Dean Karlen, Mark J. Panaggio, Matt Kinsey, Luke C. Mullany, Kaitlin Rainwater-Lovett, Lauren Shin, Katharine Tallaksen, Shelby Wilson, Michael

Brenner, Marc Coram, Jessie K. Edwards, Keya Joshi, Ellen Klein, Juan Dent Hulse, Kyra H. Grantz, Alison L. Hill, Kathryn Kaminsky, Joshua Kaminsky, Lindsay T. Keegan, Stephen A. Lauer, Elizabeth C. Lee, Joseph C. Lemaitre, Justin Lessler, Hannah R. Meredith, Javier Perez-Saez, Sam Shah, Claire P. Smith, Shaun A. Truelove, Josh Wills, Lauren Gardner, Maximilian Marshall, Kristen Nixon, John C. Burant, Jozef Budzinski, Wen-Hao Chiang, George Mohler, Junyi Gao, Lucas Glass, Cheng Qian, Justin Romberg, Rakshith Sharma, Jeffrey Spaeder, Jimeng Sun, Cao Xiao, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Yueying Wang, Guannan Wang, Lily Wang, Shan Yu, Chaman Jain, Sangeeta Bhatia, Pierre Nouvellet, Ryan Barber, Emmanuela Gaikedu, Simon Hay, Steve Lim, Chris Murray, David Pigott, Robert C. Reiner, Prasith Baccam, Heidi L. Gurung, Steven A. Stage, Bradley T. Suchoski, Chung-Yan Fong, Dit-Yan Yeung, Bijaya Adhikari, Jiaming Cui, B. Aditya Prakash, Alexander Rodríguez, Anika Tabassum, Jiajia Xie, John Asplund, Arden Baxter, Pinar Keskinocak, Buse Eylul Oruc, Nicoleta Serban, Sercan O. Arik, Mike Dusenberry, Arkady Epshteyn, Elli Kanal, Long T. Le, Chun-Liang Li, Tomas Pfister, Rajarishi Sinha, Thomas Tsai, Nate Yoder, Jinsung Yoon, Leyou Zhang, Daniel Wilson, Artur A. Belov, Carson C. Chow, Richard C. Gerkin, Osman N. Yogurtcu, Mark Ibrahim, Timothee Lacroix, Matthew Le, Jason Liao, Maximilian Nickel, Levent Sagun, Sam Abbott, Nikos I. Bosse, Sebastian Funk, Joel Hellewell, Sophie R. Meakin, Katharine Sherratt, Rahi Kalantari, Mingyuan Zhou, Morteza Karimzadeh, Benjamin Lucas, Thoai Ngo, Hamidreza Zoraghein, Behzad Vahedi, Zhongying Wang, Sen Pei, Jeffrey Shaman, Teresa K. Yamana, Dimitris Bertsimas, Michael L. Li, Saksham Soni, Hamza Tazi Bouardi, Madeline Adee, Turgay Ayer, Jagpreet Chhatwal, Ozden O. Dalgic, Mary A. Ladd, Benjamin P. Linas, Peter Mueller, Jade Xiao, Jurgen Bosch, Austin Wilson, Peter Zimmerman, Qinxia Wang, Yuanjia Wang, Shanghong Xie, Donglin Zeng, Jacob Bien, Logan Brooks, Alden Green, Addison J. Hu, Maria Jahja, Daniel McDonald, Balasubramanian Narasimhan, Collin Politsch, Samyak Rajanala, Aaron Rumack, Noah Simon, Ryan J. Tibshirani, Rob Tibshirani, Valerie Ventura, Larry Wasserman, John M. Drake, Eamon B. O'Dea, Yaser Abu-Mostafa, Rahil Bathwal, Nicholas A. Chang, Pavan Chitta, Anne Erickson, Sumit Goel, Jethin Gowda, Qixuan Jin, HyeongChan Jo, Juhyun Kim, Pranav Kulkarni, Samuel M. Lushtak, Ethan Mann, Max Popken, Connor Soohoo, Kushal Tirumala, Albert Tseng, Vignesh Varadarajan, Jagath Vytheeswaran, Christopher Wang, Akshay Yeluri, Dominic Yurk, Michael Zhang, Alexander Zlokapa, Robert Pagano, Chandini Jain, Vishal Tomar, Lam Ho, Huong Huynh, Quoc Tran, Velma K. Lopez, Jo W. Walker, Rachel B. Slayton, Michael A. Johansson, Matthew Biggerstaff, and Nicholas G. Reich. The United States COVID-19 Forecast Hub dataset. *Scientific Data*, 9(1):462, August 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01517-w. URL https://www.nature.com/articles/s41597-022-01517-w.

[174] Adam J. Kucharski, Emma B. Hodcroft, and Moritz U. G. Kraemer. Sharing, synthesis and sustainability of data analysis for epidemic preparedness in Europe. *The Lancet Regional Health – Europe*, 9, October 2021. ISSN 2666-7762. doi: 10.1016/j.lanepe.2021.100215. URL https://www.thelancet.com/journals/lanepe/article/PIIS2666-7762(21)00192-7/fulltext. Publisher: Elsevier.

[175] Freya M. Shearer and Marc Lipsitch. The importance of playing the long game when it comes to pandemic surveillance. *Proceedings of the National Academy of Sciences*, 122(15):e2500328122, April 2025. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2500328122. URL https://pnas.org/doi/10.1073/pnas.2500328122.

[176] J. Clarke, A. Lim, P. Gupte, D. M. Pigott, W. G. van Panhuis, and O. J. Brady. A global dataset of publicly available dengue case count data. *Scientific Data*, 11(1):296, March 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-03120-7. URL https://www.nature.com/articles/s41597-024-03120-7. Publisher: Nature Publishing Group.

[177] S. Runge-Ranzinger, P. J. McCall, A. Kroeger, and O. Horstick. Dengue disease surveillance: an updated systematic literature review. *Tropical Medicine & International Health*, 19(9):1116–1160, September 2014. ISSN 1360-2276, 1365-3156. doi: 10.1111/tmi.12333. URL https://onlinelibrary.wiley.com/doi/10.1111/tmi.12333.

[178] Eduardo A. Undurraga, Yara A. Halasa, and Donald S. Shepard. Use of Expansion Factors to Estimate the Burden of Dengue in Southeast Asia: A Systematic Analysis. *PLOS Neglected Tropical Diseases*, 7(2):e2056, February 2013. ISSN 1935-2735. doi: 10.1371/journal.pntd.0002056. URL https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002056. Publisher: Public Library of Science.

[179] Monaise M.O. Silva, Moreno S. Rodrigues, Igor A.D. Paploski, Mariana Kikuti, Amelia M. Kasper, Jaqueline S. Cruz, Tássia L. Queiroz, Aline S. Tavares, Perla M. Santana, Josélio M.G. Araújo, Albert I. Ko, Mitermayer Galvão Reis, and Guilherme S. Ribeiro. Accuracy of Dengue Reporting by National Surveillance System, Brazil. *Emerging Infectious Diseases*, 22(2):336–339, February 2016. ISSN 1080-6040, 1080-6059. doi: 10.3201/eid2202.150495. URL http://wwwnc.cdc.gov/eid/article/22/2/15-0495_article.htm.

[180] Amitabh B. Suthar, Sara Schubert, Julie Garon, Alexia Couture, Amy M. Brown, and Sana Charania. Coronavirus Disease Case Definitions, Diagnostic Testing Criteria, and Surveillance in 25 Countries with Highest Reported Case Counts. *Emerging Infectious Diseases*, 28(1):148–156, January 2022. ISSN 1080-6040. doi: 10.3201/eid2801.211082. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8714223/.

[181] Tim K. Tsang, Peng Wu, Yun Lin, Eric H. Y. Lau, Gabriel M. Leung, and Benjamin J. Cowling. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *The Lancet Public Health*, 5(5):e289–e296, May 2020. ISSN 2468-2667. doi: 10.1016/S2468-2667(20)30089-X. URL https://www.thelancet.com/journals/lanpub/article/PIIS2468-26672030089-X/fulltext. Publisher: Elsevier.

[182] World Health Organization. WHO COVID-19 Case Definition 2020, July 2022. URL https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2022.1.

[183] Cheryl L Gibbons, Marie-Josée J Mangen, Dietrich Plass, Arie H Havelaar, Russell John Brooke, Piotr Kramarz, Karen L Peterson, Anke L Stuurman, Alessandro Cassini, Eric M Fèvre, and Mirjam E.E. Kretzschmar. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health*, 14(1):147, December 2014. ISSN 1471-2458. doi: 10.1186/1471-2458-14-147. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-14-147.

[184] Lilit Kazazian, Antonio S. Lima Neto, Geziel S. Sousa, Osmar José Do Nascimento, and Marcia C. Castro. Spatiotemporal transmission dynamics of co-circulating dengue, Zika, and chikungunya viruses in Fortaleza, Brazil: 2011–2017. *PLOS Neglected Tropical Diseases*, 14(10):e0008760, October 2020. ISSN 1935-2735. doi: 10.1371/journal.pntd.0008760. URL https://dx.plos.org/10.1371/journal.pntd.0008760.

[185] Cristiane W. Cardoso, Igor A. D. Paploski, Mariana Kikuti, Moreno S. Rodrigues, Monaise M. O. Silva, Gubio S. Campos, Silvia I. Sardi, Uriel Kitron, Mitermayer G. Reis, and Guilherme S. Ribeiro. Outbreak of Exanthematous Illness Associated with Zika, Chikungunya, and Dengue Viruses, Salvador, Brazil - Volume 21, Number 12—December 2015 - Emerging Infectious Diseases journal - CDC. December 2015. doi: 10.3201/eid2112.151167. URL https://wwwnc.cdc.gov/eid/article/21/12/15-1167_article.

[186] D. J. Weiss, A. Nelson, C. A. Vargas-Ruiz, K. Gligorić, S. Bavadekar, E. Gabrilovich, A. Bertozzi-Villa, J. Rozier, H. S. Gibson, T. Shekel, C. Kamath, A. Lieber, K. Schulman, Y. Shao, V. Qarkaxhija, A. K. Nandi, S. H. Keddie, S. Rumisha, P. Amratia, R. Arambepola, E. G. Chestnutt, J. J. Millar, T. L. Symons, E. Cameron, K. E. Battle, S. Bhatt, and P. W. Gething. Global maps of travel time to healthcare facilities. *Nature Medicine*, 26(12):1835–1838, December 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-1059-1. URL https://www.nature.com/articles/s41591-020-1059-1. Publisher: Nature Publishing Group.

[187] Casey M. Zipfel, Vittoria Colizza, and Shweta Bansal. Health inequities in influenza transmission and surveillance. *PLOS Computational Biology*, 17(3):e1008642, March 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008642. URL https://dx.plos.org/10.1371/journal.pcbi.1008642.

[188] Judith A. Bouman, Anthony Hauser, Simon L. Grimm, Martin Wohlfender, Samir Bhatt, Elizaveta Semenova, Andrew Gelman, Christian L. Althaus, and Julien Riou. Bayesian workflow for time-varying transmission in stratified compartmental infectious disease transmission models. *PLOS Computational Biology*, 20(4):e1011575, April 2024. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011575. URL https://dx.plos.org/10.1371/journal.pcbi.1011575.

[189] James A. Hay, Isobel Routledge, and Saki Takahashi. Serodynamics: A primer and synthetic review of methods for epidemiological inference using serological data. *Epidemics*, 49:100806, December 2024. ISSN 17554365. doi: 10.1016/j.epidem.2024.100806. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436524000677.

[190] Henrik Salje, Justin Lessler, Irina Maljkovic Berry, Melanie C. Melendrez, Timothy Endy, Siripen Kalayanarooj, Atchareeya A-Nuegoonpipat, Sumalee Chanama, Somchai Sangkijporn, Chonticha Klungthong, Butsaya Thaisomboonsuk, Ananda Nisalak, Robert V. Gibbons, Sopon Iamsirithaworn, Louis R. Macareo, In-Kyu Yoon, Areerat Sangarsang, Richard G. Jarman, and Derek A. T. Cummings. Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. *Science*, 355(6331):1302–1306, March 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaj9384. URL https://www.science.org/doi/10.1126/science.aaj9384.

[191] Heidi E. Brown, Luigi Sedda, Chris Sumner, Elene Stefanakos, Irene Ruberto, and Matthew Roach. Understanding Mosquito Surveillance Data for Analytic Efforts: A Case Study. *Journal of Medical Entomology*, 58(4):1619–1625, July 2021. ISSN 1938-2928. doi: 10.1093/jme/tjab018.

[192] Kris V. Parag, Ben Lambert, Christl A. Donnelly, and Sandor Beregi. Asymmetric limits on timely interventions from noisy epidemic data, January 2025. URL https://www.medrxiv.org/content/10.1101/2025.01.28.25321278v1. Pages: 2025.01.28.25321278.

[193] Bethan Cracknell Daniels, Neil Ferguson, and Ilaria Dorigatti. Efficacy, public health impact and optimal use of the Takeda dengue vaccine, August 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.08.10.24311393.

[194] Geoffrey Rose. Sick Individuals and Sick Populations. *International Journal of Epidemiology*, 14(1):32–38, 1985. ISSN 0300-5771, 1464-3685. doi: 10.1093/ije/14.1.32. URL https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/14.1.32.

[195] M. U. G. Kraemer, T. A. Perkins, D. a. T. Cummings, R. Zakar, S. I. Hay, D. L. Smith, and R. C. Reiner. Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *Journal of The Royal Society Interface*, 12(111):20150468, October 2015. doi: 10.1098/rsif.2015.0468. URL https://royalsocietypublishing.org/doi/full/10.1098/rsif.2015.0468. Publisher: Royal Society.

[196] Paul Reiter. Climate change and mosquito-borne disease. *Environmental Health Perspectives*, 109(suppl 1):141–161, March 2001. ISSN 0091-6765, 1552-9924. doi: 10.1289/ehp.01109s1141. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.01109s1141.

[197] Katherine Mackey, Chelsea K. Ayers, Karli K. Kondo, Somnath Saha, Shailesh M. Advani, Sarah Young, Hunter Spencer, Max Rusek, Johanna Anderson, Stephanie Veazie, Mia Smith, and Devan Kansagara. Racial and Ethnic Disparities in COVID-19–Related Infections, Hospitalizations, and Deaths: A Systematic Review. *Annals of Internal Medicine*, 174(3):362–373, March 2021. ISSN 0003-4819, 1539-3704. doi: 10.7326/M20-6306. URL https://www.acpjournals.org/doi/10.7326/M20-6306.

[198] Robert C. Reiner, T. Alex Perkins, Christopher M. Barker, Tianchan Niu, Luis Fernando Chaves, Alicia M. Ellis, Dylan B. George, Arnaud Le Menach, Juliet R. C. Pulliam, Donal Bisanzio, Caroline Buckee, Christinah Chiyaka, Derek A. T. Cummings, Andres J. Garcia, Michelle L. Gatton, Peter W. Gething, David M. Hartley, Geoffrey Johnston, Eili Y. Klein, Edwin Michael, Steven W. Lindsay, Alun L. Lloyd, David M. Pigott, William K. Reisen, Nick Ruktanonchai, Brajendra K. Singh, Andrew J. Tatem, Uriel Kitron, Simon I. Hay, Thomas W. Scott, and David L. Smith. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *Journal of The Royal Society Interface*, 10(81):20120921, April 2013. doi: 10.1098/rsif.2012.0921. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2012.0921. Publisher: Royal Society.

[199] Rosalind M. Eggo, Jeanette Dawa, Adam J. Kucharski, and Zulma M. Cucunuba. The importance of local context in COVID-19 models. *Nature Computational Science*, 1(1):6–8, January 2021. ISSN 2662-8457. doi: 10.1038/s43588-020-00014-7. URL https://www.nature.com/articles/s43588-020-00014-7. Publisher: Nature Publishing Group.

[200] Sophie A. Lee, Christopher I. Jarvis, W. John Edmunds, Theodoros Economou, and Rachel Lowe. Spatial connectivity in mosquito-borne disease models: a systematic review of methods and assumptions. *Journal of The Royal Society Interface*, 18(178):20210096, May 2021. doi: 10.1098/rsif.2021.0096. URL https://royalsocietypublishing.org/doi/full/10.1098/rsif.2021.0096. Publisher: Royal Society.

[201] Jing Li, Daniel Blakeley, and Robert J. Smith. The failure of R0. *Computational and Mathematical Methods in Medicine*, 2011:527610, 2011. ISSN 1748-6718. doi: 10.1155/2011/527610.

[202] Danielle M Richard and Marc Lipsitch. What's next: using infectious disease mathematical modelling to address health disparities. *International Journal of Epidemiology*, 53(1):dyad180, February 2024. ISSN 0300-5771, 1464-3685. doi: 10.1093/ije/dyad180. URL https://academic.oup.com/ije/article/doi/10.1093/ije/dyad180/7497080.

[203] Ruth McCabe, Mara D. Kont, Nora Schmit, Charles Whittaker, Alessandra Løchen, Patrick G.T. Walker, Azra C. Ghani, Neil M. Ferguson, Peter J. White, Christl A. Donnelly, and Oliver J. Watson. Communicating uncertainty in epidemic models. *Epidemics*, 37:100520, December 2021. ISSN 17554365. doi: 10.1016/j.epidem.2021.100520. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436521000669.

[204] Matthew J. Penn, Daniel J. Laydon, Joseph Penn, Charles Whittaker, Christian Morgenstern, Oliver Ratmann, Swapnil Mishra, Mikko S. Pakkanen, Christl A. Donnelly, and Samir Bhatt. Intrinsic randomness in epidemic modelling beyond statistical uncertainty. *Communications Physics*, 6(1):146, June 2023. ISSN 2399-3650. doi: 10.1038/s42005-023-01265-2. URL https://www.nature.com/articles/s42005-023-01265-2.

[205] Liza Hadley, Caylyn Rich, Alex Tasker, Olivier Restif, and Sebastian Funk. Visual preferences for communicating modelling: a global analysis of COVID-19 policy and decision makers. *Infectious Disease Modelling*, 10(3):924–934, September 2025. ISSN 24680427. doi: 10.1016/j.idm.2025.04.005. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042725000314.

[206] Yoshitsugu Oono. *The Nonlinear World: Conceptual Analysis and Phenomenology*. Springer Series in Synergetics. Springer Japan, Tokyo, 2013. ISBN 978-4-431-54028-1 978-4-431-54029-8. doi: 10.1007/978-4-431-54029-8. URL https://link.springer.com/10.1007/978-4-431-54029-8.

[207] Steven Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. A Chapman & Hall book. CRC Press, Boca Raton London New York, second edition, first issued in hardback edition, 2019. ISBN 978-0-8133-4910-7 978-0-367-09206-1.

[208] C. J. Rhodes and R. M. Anderson. Power laws governing epidemics in isolated populations. *Nature*, 381(6583): 600–602, June 1996. ISSN 0028-0836. doi: 10.1038/381600a0.

[209] Adam J. Kucharski and Marc Baguelin. The role of human immunity and social behavior in shaping influenza evolution. *PLOS Pathogens*, 13(8):e1006432, August 2017. ISSN 1553-7374. doi: 10.1371/journal.ppat.1006432. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1006432. Publisher: Public Library of Science.

[210] John L. Spouge. A comprehensive estimation of country-level basic reproduction numbers R0 for COVID-19: Regime regression can automatically estimate the end of the exponential phase in epidemic data. *PLOS ONE*, 16(7): e0254145, July 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0254145. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0254145. Publisher: Public Library of Science.

[211] K. Dietz. The Incidence of Infectious Diseases under the Influence of Seasonal Fluctuations. In S. Levin, Jürgen Berger, Wolfgang J. Bühler, Rudolf Repges, and Petre Tautu, editors, *Mathematical Models in Medicine*, volume 11, pages 1–15. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976. ISBN 978-3-540-07802-9 978-3-642-93048-5. doi: 10.1007/978-3-642-93048-5_1. URL http://link.springer.com/10.1007/978-3-642-93048-5_1. Series Title: Lecture Notes in Biomathematics.

[212] Stephen P. Ellner and John Guckenheimer. *Dynamic Models in Biology*. Princeton University Press, September 2011. ISBN 978-1-4008-4096-0 978-0-691-12589-3. doi: 10.2307/j.ctvcm4h1q. URL http://www.jstor.org/stable/10.2307/j.ctvcm4h1q.

[213] Nikos I. Bosse, Sam Abbott, Anne Cori, Edwin Van Leeuwen, Johannes Bracher, and Sebastian Funk. Scoring epidemiological forecasts on transformed scales. *PLOS Computational Biology*, 19(8):e1011393, August 2023. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011393. URL https://dx.plos.org/10.1371/journal.pcbi.1011393.

[214] Rachel Lowe, Antonio Gasparrini, Cédric J. Van Meerbeeck, Catherine A. Lippi, Roché Mahon, Adrian R. Trotman, Leslie Rollock, Avery Q. J. Hinds, Sadie J. Ryan, and Anna M. Stewart-Ibarra. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLOS Medicine*, 15(7):e1002613, July 2018. ISSN 1549-1676. doi: 10.1371/journal.pmed.1002613. URL https://dx.plos.org/10.1371/journal.pmed.1002613.

[215] Louis Lambrechts, Krijn P. Paaijmans, Thanyalak Fansiri, Lauren B. Carrington, Laura D. Kramer, Matthew B. Thomas, and Thomas W. Scott. Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti. *Proceedings of the National Academy of Sciences*, 108(18):7460–7465, May 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1101377108. URL https://pnas.org/doi/full/10.1073/pnas.1101377108.

[216] Isaac J. Stopard, Antoine Sanou, Eunho Suh, Lauren J. Cator, Matthew B. Thomas, W. Moussa Guelbéogo, N'Falé Sagnon, Ben Lambert, and Thomas S. Churcher. Modelling the effects of diurnal temperature variation on malaria infection dynamics in mosquitoes. *Communications Biology*, 8(1):1–13, April 2025. ISSN 2399-3642. doi: 10.1038/s42003-025-07949-5. URL https://www.nature.com/articles/s42003-025-07949-5. Publisher: Nature Publishing Group.

[217] H. M. Yang, M. L. G. Macoris, K. C. Galvani, M. T. M. Andrighetti, and D. M. V. Wanderley. Assessing the effects of temperature on the population of Aedes aegypti , the vector of dengue. *Epidemiology and Infection*, 137(8):1188–1202, August 2009. ISSN 0950-2688, 1469-4409. doi: 10.1017/S0950268809002040. URL https://www.cambridge.org/core/product/identifier/S0950268809002040/type/journal_article.

[218] T. Alex Perkins, Thomas W. Scott, Arnaud Le Menach, and David L. Smith. Heterogeneity, Mixing, and the Spatial Scales of Mosquito-Borne Pathogen Transmission. *PLoS Computational Biology*, 9(12):e1003327, December 2013. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003327. URL https://dx.plos.org/10.1371/journal.pcbi.1003327.

[219] Devin Kirk, Samantha Straus, Marissa L. Childs, Mallory Harris, Lisa Couper, T. Jonathan Davies, Coreen Forbes, Alyssa-Lois Gehman, Maya L. Groner, Christopher Harley, Kevin D. Lafferty, Van Savage, Eloise Skinner, Mary O'Connor, and Erin A. Mordecai. Temperature impacts on dengue incidence are nonlinear and mediated by climatic and socioeconomic factors: A meta-analysis. *PLOS Climate*, 3(3):e0000152, March 2024. ISSN 2767-3200. doi: 10.1371/journal.pclm.0000152. URL https://dx.plos.org/10.1371/journal.pclm.0000152.

[220] Alasdair D. Henderson, Mike Kama, Maite Aubry, Stephane Hue, Anita Teissier, Taina Naivalu, Vinaisi D. Bechu, Jimaima Kailawadoko, Isireli Rabukawaqa, Aalisha Sahukhan, Martin L. Hibberd, Eric J. Nilles, Sebastian Funk, Jimmy Whitworth, Conall H. Watson, Colleen L. Lau, W. John Edmunds, Van-Mai Cao-Lormeau, and Adam J. Kucharski. Interactions between timing and transmissibility explain diverse flavivirus dynamics in Fiji. *Nature Communications*, 12(1):1671, March 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21788-y. URL https://www.nature.com/articles/s41467-021-21788-y. Publisher: Nature Publishing Group.

[221] Duncan J. Watts, Roby Muhamad, Daniel C. Medina, and Peter S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences*, 102(32):11157–11162, August 2005. doi: 10.1073/pnas.0501226102. URL https://www.pnas.org/doi/10.1073/pnas.0501226102. Publisher: Proceedings of the National Academy of Sciences.

[222] Mokhalad A. Majeed, Helmi Zulhaidi Mohd Shafri, Zed Zulkafli, and Aimrun Wayayok. A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using LSTM with Spatial Attention. *International Journal of Environmental Research and Public Health*, 20(5):4130, February 2023. ISSN 1660-4601. doi: 10.3390/ijerph20054130. URL https://www.mdpi.com/1660-4601/20/5/4130.

[223] Nicolas Banholzer, Thomas Mellan, H. Juliette T. Unwin, Stefan Feuerriegel, Swapnil Mishra, and Samir Bhatt. A comparison of short-term probabilistic forecasts for the incidence of COVID-19 using mechanistic and statistical time series models, May 2023. URL http://arxiv.org/abs/2305.00933. arXiv:2305.00933 [cs, q-bio, stat].

[224] M. Naser Lessani, Zhenlong Li, Fengrui Jing, Shan Qiao, Jiajia Zhang, Bankole Olatosi, and Xiaoming Li. Human mobility and the infectious disease transmission: a systematic review. *Geo-spatial Information Science*, 27(6): 1824–1851, November 2024. ISSN 1009-5020, 1993-5153. doi: 10.1080/10095020.2023.2275619. URL https://www.tandfonline.com/doi/full/10.1080/10095020.2023.2275619.

[225] M. U. G. Kraemer, N. Golding, D. Bisanzio, S. Bhatt, D. M. Pigott, S. E. Ray, O. J. Brady, J. S. Brownstein, N. R. Faria, D. A. T. Cummings, O. G. Pybus, D. L. Smith, A. J. Tatem, S. I. Hay, and R. C. Reiner. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Scientific Reports*, 9(1):5151, March 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-41192-3. URL https://www.nature.com/articles/s41598-019-41192-3.

[226] Suttikiat Changruenngam, Dominique J. Bicout, and Charin Modchang. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. *Scientific Reports*, 10(1):11325, July 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-68230-9. URL https://www.nature.com/articles/s41598-020-68230-9.

[227] Vinyas Harish, Felipe J. Colón-González, Filipe R. R. Moreira, Rory Gibb, Moritz U. G. Kraemer, Megan Davis, Robert C. Reiner, David M. Pigott, T. Alex Perkins, Daniel J. Weiss, Isaac I. Bogoch, Gonzalo Vazquez-Prokopec,

Pablo Manrique Saide, Gerson L. Barbosa, Ester C. Sabino, Kamran Khan, Nuno R. Faria, Simon I. Hay, Fabián Correa-Morales, Francisco Chiaravalloti-Neto, and Oliver J. Brady. Human movement and environmental barriers shape the emergence of dengue. *Nature Communications*, 15(1):4205, May 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-48465-0. URL https://www.nature.com/articles/s41467-024-48465-0. Publisher: Nature Publishing Group.

[228] Anuj Srivastava and Gerardo Chowell. Understanding Spatial Heterogeneity of COVID-19 Pandemic Using Shape Analysis of Growth Rate Curves, May 2020. URL http://medrxiv.org/lookup/doi/10.1101/2020.05.25.20112433.

[229] Magdalene Walters and T. Alex Perkins. Hidden heterogeneity and its influence on dengue vaccination impact. *Infectious Disease Modelling*, 5:783–797, January 2020. ISSN 2468-0427. doi: 10.1016/j.idm.2020.09.008. URL https://www.sciencedirect.com/science/article/pii/S2468042720300531.

[230] Benjamin D. Dalziel, Stephen Kissler, Julia R. Gog, Cecile Viboud, Ottar N. Bjørnstad, C. Jessica E. Metcalf, and Bryan T. Grenfell. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science*, 362(6410):75–79, October 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aat6030. URL https://www.science.org/doi/10.1126/science.aat6030.

[231] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, December 2009. doi: 10.1073/pnas.0906910106. URL https://www.pnas.org/doi/10.1073/pnas.0906910106. Publisher: Proceedings of the National Academy of Sciences.

[232] Laura Andrea Barrero Guevara, Sarah C. Kramer, Tobias Kurth, and Matthieu Domenech De Cellès. Causal inference concepts can guide research into the effects of climate on infectious diseases. *Nature Ecology & Evolution*, 9(2):349–363, November 2024. ISSN 2397-334X. doi: 10.1038/s41559-024-02594-3. URL https://www.nature.com/articles/s41559-024-02594-3.

[233] Jorge L. B. Araújo, Rafael Bomfim, Cesar I. N. Sampaio Filho, Luciano P. G. Cavalcanti, Antonio S. Lima Neto, José S. Andrade, and Vasco Furtado. The impact of COVID-19 mobility restrictions on dengue transmission in urban areas. *PLOS Neglected Tropical Diseases*, 18(11):e0012644, November 2024. ISSN 1935-2727. doi: 10.1371/journal.pntd.0012644. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11627415/.

[234] Luca Ferretti, Chris Wymant, James Petrie, Daphne Tsallis, Michelle Kendall, Alice Ledda, Francesco Di Lauro, Adam Fowler, Andrea Di Francia, Jasmina Panovska-Griffiths, Lucie Abeler-Dörner, Marcos Charalambides, Mark Briers, and Christophe Fraser. Digital measurement of SARS-CoV-2 transmission risk from 7 million contacts. *Nature*, 626(7997):145–150, February 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-023-06952-2. URL https://www.nature.com/articles/s41586-023-06952-2.

[235] Qinghui Zeng, Xiaolin Yu, Haobo Ni, Lina Xiao, Ting Xu, Haisheng Wu, Yuliang Chen, Hui Deng, Yingtao Zhang, Sen Pei, Jianpeng Xiao, and Pi Guo. Dengue transmission dynamics prediction by combining metapopulation networks and Kalman filter algorithm. *PLOS Neglected Tropical Diseases*, 17(6):e0011418, June 2023. ISSN 1935-2735. doi: 10.1371/journal.pntd.0011418. URL https://dx.plos.org/10.1371/journal.pntd.0011418.

[236] Angkana T. Huang, Saki Takahashi, Henrik Salje, Lin Wang, Bernardo Garcia-Carreras, Kathryn Anderson, Timothy Endy, Stephen Thomas, Alan L. Rothman, Chonticha Klungthong, Anthony R. Jones, Stefan Fernandez, Sopon Iamsirithaworn, Pawinee Doung-Ngern, Isabel Rodriguez-Barraquer, and Derek A. T. Cummings. Assessing the role of multiple mechanisms increasing the age of dengue cases in Thailand. *Proceedings of the National Academy of*

Sciences, 119(20):e2115790119, May 2022. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2115790119. URL https://pnas.org/doi/full/10.1073/pnas.2115790119.

[237] Paula Moraga. *Geospatial health data: modeling and visualization with R-INLA and Shiny*. Chapman & Hall/CRC biostatistics series. CRC Press/Taylor & Francis Group, Boca Raton, 2020. ISBN 978-0-367-35795-5.

[238] Jack Wardle, Sangeeta Bhatia, Moritz U.G. Kraemer, Pierre Nouvellet, and Anne Cori. Gaps in mobility data and implications for modelling epidemic spread: A scoping review and simulation study. *Epidemics*, 42:100666, March 2023. ISSN 17554365. doi: 10.1016/j.epidem.2023.100666. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436523000026.

[239] C. E. Gehlke and Katherine Biehl. Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material. *Journal of the American Statistical Association*, 29(185):169, March 1934. ISSN 01621459. doi: 10.2307/2277827. URL https://www.jstor.org/stable/2277827?origin=crossref.

[240] Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 0 edition, September 2014. ISBN 978-0-429-13717-4. doi: 10.1201/b17115. URL https://www.taylorfrancis.com/books/9781439819180.

[241] Hyunju Lee, Heeyoung Lee, Kyoung-Ho Song, Eu Suk Kim, Jeong Su Park, Jongtak Jung, Soyeon Ahn, Eun Kyeong Jeong, Hyekyung Park, and Hong Bin Kim. Impact of Public Health Interventions on Seasonal Influenza Activity During the COVID-19 Outbreak in Korea. *Clinical Infectious Diseases*, 73(1):e132–e140, July 2021. ISSN 1058-4838, 1537-6591. doi: 10.1093/cid/ciaa672. URL https://academic.oup.com/cid/article/73/1/e132/5848838.

[242] Andrew J. Tatem. WorldPop, open data for spatial demography. *Scientific Data*, 4(1):170004, January 2017. ISSN 2052-4463. doi: 10.1038/sdata.2017.4. URL https://www.nature.com/articles/sdata20174. Publisher: Nature Publishing Group.

[243] Jennifer Beam Dowd, Liliana Andriano, David M. Brazel, Valentina Rotondi, Per Block, Xuejie Ding, Yan Liu, and Melinda C. Mills. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences*, 117(18):9696–9698, May 2020. doi: 10.1073/pnas.2004911117. URL https://www.pnas.org/doi/abs/10.1073/pnas.2004911117. Publisher: Proceedings of the National Academy of Sciences.

[244] Isabel Rodriguez-Barraquer, Henrik Salje, and Derek A Cummings. Opportunities for improved surveillance and control of dengue from age-specific case data. *eLife*, 8:e45474, May 2019. ISSN 2050-084X. doi: 10.7554/eLife.45474. URL https://elifesciences.org/articles/45474.

[245] James D. Munday, Sam Abbott, Sophie Meakin, and Sebastian Funk. Evaluating the use of social contact data to produce age-specific short-term forecasts of SARS-CoV-2 incidence in England. *PLOS Computational Biology*, 19(9): e1011453, September 2023. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011453. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011453. Publisher: Public Library of Science.

[246] Rachel Lowe and Claudia Torres Codeço. Harmonizing Multisource Data to Inform Vector-Borne Disease Risk Management Strategies. *Annual Review of Entomology*, October 2024. ISSN 0066-4170, 1545-4487. doi: 10.1146/annurev-ento-040124-015101. URL https://www.annualreviews.org/content/journals/10.1146/annurev-ento-040124-015101.

[247] Shweta Bansal, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, and Cécile Viboud. Big Data for Infectious Disease Surveillance and Modeling. *Journal of Infectious Diseases*, 214(suppl 4):S375–S379, December 2016. ISSN 0022-1899, 1537-6613. doi: 10.1093/infdis/jiw400. URL https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiw400.

[248] Robin N. Thompson and Ellen Brooks-Pollock. Detection, forecasting and control of infectious disease epidemics: modelling outbreaks in humans, animals and plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1775):20190038, June 2019. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2019.0038. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0038.

[249] Xiao-Li Meng. Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, 12(2):685–726, June 2018. ISSN 1932-6157, 1941-7330. doi: 10.1214/18-AOAS1161SF. URL https://projecteuclid.org/journals/annals-of-applied-statistics/volume-12/issue-2/Statistical-paradises-and-paradoxes-in-big-data-I--Law/10.1214/18-AOAS1161SF.full. Publisher: Institute of Mathematical Statistics.

[250] Amanda J. Meadows, Ben Oppenheim, Jaclyn Guerrero, Benjamin Ash, Rinette Badker, Cathine K. Lam, Chris Pardee, Christopher Ngoon, Patrick T. Savage, Vikram Sridharan, Nita K. Madhav, and Nicole Stephenson. Infectious Disease Underreporting Is Predicted by Country-Level Preparedness, Politics, and Pathogen Severity. *Health Security*, 20(4):331–338, August 2022. ISSN 2326-5094, 2326-5108. doi: 10.1089/hs.2021.0197. URL https://www.liebertpub.com/doi/10.1089/hs.2021.0197.

[251] Emmanuelle A. Dankwa, Andrew F. Brouwer, and Christl A. Donnelly. Structural identifiability of compartmental models for infectious disease transmission is influenced by data type. *Epidemics*, 41:100643, December 2022. ISSN 17554365. doi: 10.1016/j.epidem.2022.100643. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436522000834.

[252] Luca Gallo, Mattia Frasca, Vito Latora, and Giovanni Russo. Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models. *Science Advances*, 8(3):eabg5234, January 2022. ISSN 2375-2548. doi: 10.1126/sciadv.abg5234. URL https://www.science.org/doi/10.1126/sciadv.abg5234.

[253] Natsuko Imai, Ilaria Dorigatti, Anne Cori, Steven Riley, and Neil Ferguson. Report 1: Estimating the potential total number of novel Coronavirus cases in Wuhan City, China. January 2020. doi: 10.25561/77149. URL http://hdl.handle.net/10044/1/77149.

[254] Nicole A. Scavo, Chalmers Vasquez, John-Paul Mutebi, Marco Ajelli, and André B. B. Wilke. Leveraging mosquito surveillance data to assess ecosystem health. *Trends in Parasitology*, May 2025. ISSN 1471-4922. doi: 10.1016/j.pt.2025.04.011. URL https://www.sciencedirect.com/science/article/pii/S1471492225001059.

[255] Catherine A Lippi, Samuel S C Rund, and Sadie J Ryan. Characterizing the Vector Data Ecosystem. *Journal of Medical Entomology*, 60(2):247–254, February 2023. ISSN 0022-2585. doi: 10.1093/jme/tjad009. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9989832/.

[256] Imelda K. Moise, Leo C. Zulu, Douglas O. Fuller, and John C. Beier. Persistent Barriers to Implementing Efficacious Mosquito Control Activities in the Continental United States: Insights from Vector Control Experts. In Alfonso J. Rodriguez-Morales, editor, *Current Topics in Neglected Tropical Diseases*. IntechOpen, December 2019. ISBN 978-1-78923-889-1 978-1-78923-890-7. doi: 10.5772/intechopen.

76774. URL https://www.intechopen.com/books/current-topics-in-neglected-tropical-diseases/persistent-barriers-to-implementing-efficacious-mosquito-control-activities-in-the-continental-unite.

[257] G. R. William Wint, Thomas Balenghien, Eduardo Berriatua, Marieta Braks, Cedric Marsboom, Jolyon Medlock, Francis Schaffner, Wim Van Bortel, Neil Alexander, Bulent Alten, Ewelina Czwienczek, Sofie Dhollander, Els Ducheyne, Celine M. Gossner, Kayleigh Hansford, Guy Hendrickx, Hector Honrubia, Tom Matheussen, Andrei Daniel Mihalca, Dusan Petric, Jane Richardson, Hein Sprong, Veerle Versteirt, and Olivier Briet. VectorNet: collaborative mapping of arthropod disease vectors in Europe and surrounding areas since 2010. *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, 28(26):2200666, June 2023. ISSN 1560-7917. doi: 10.2807/1560-7917.ES.2023.28.26.2200666.

[258] Ah-Young Lim, Yalda Jafari, Jamie M. Caldwell, Hannah E. Clapham, Katy A. M. Gaythorpe, Laith Hussain-Alkhateeb, Michael A. Johansson, Moritz U. G. Kraemer, Richard J. Maude, Clare P. McCormack, Jane P. Messina, Erin A. Mordecai, Ingrid B. Rabe, Robert C. Reiner, Sadie J. Ryan, Henrik Salje, Jan C. Semenza, Diana P. Rojas, and Oliver J. Brady. A systematic review of the data, methods and environmental covariates used to map Aedes-borne arbovirus transmission risk. *BMC Infectious Diseases*, 23(1):708, October 2023. ISSN 1471-2334. doi: 10.1186/s12879-023-08717-8. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-023-08717-8.

[259] Abdollah Jalilian, Jorge Mateu, and Luigi Sedda. A brief review and guidance on the spatiotemporal sampling designs for disease vector surveillance. *Current Research in Parasitology & Vector-Borne Diseases*, 6:100208, January 2024. ISSN 2667-114X. doi: 10.1016/j.crpvbd.2024.100208. URL https://www.sciencedirect.com/science/article/pii/S2667114X24000396.

[260] Moritz Ug Kraemer, Marianne E Sinka, Kirsten A Duda, Adrian Qn Mylne, Freya M Shearer, Christopher M Barker, Chester G Moore, Roberta G Carvalho, Giovanini E Coelho, Wim Van Bortel, Guy Hendrickx, Francis Schaffner, Iqbal Rf Elyazar, Hwa-Jen Teng, Oliver J Brady, Jane P Messina, David M Pigott, Thomas W Scott, David L Smith, Gr William Wint, Nick Golding, and Simon I Hay. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife*, 4:e08347, June 2015. ISSN 2050-084X. doi: 10.7554/eLife.08347. URL https://elifesciences.org/articles/08347.

[261] Eleonora Cella, Mónica V. Cunha, Luiz Carlos Junior Alcantara, José Lourenço, and Marta Giovanetti. Strengthening arbovirus surveillance: bridging gaps for global health preparedness. *Trends in Microbiology*, 0(0), April 2025. ISSN 0966-842X, 1878-4380. doi: 10.1016/j.tim.2025.03.010. URL https://www.cell.com/trends/microbiology/abstract/S0966-842X(25)00106-4. Publisher: Elsevier.

[262] Simon Pollett, Benjamin M. Althouse, Brett Forshey, George W. Rutherford, and Richard G. Jarman. Internet-based biosurveillance methods for vector-borne diseases: Are they novel public health tools or just novelties? *PLOS Neglected Tropical Diseases*, 11(11):e0005871, November 2017. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005871. URL https://dx.plos.org/10.1371/journal.pntd.0005871.

[263] Beniamino Caputo, Mattia Manica, Federico Filipponi, Marta Blangiardo, Pietro Cobre, Luca Delucchi, Carlo Maria De Marco, Luca Iesu, Paola Morano, Valeria Petrella, Marco Salvemini, Cesare Bianchi, and Alessandra Della Torre. ZanzaMapp: A Scalable Citizen Science Tool to Monitor Perception of Mosquito Abundance and Nuisance in Italy and Beyond. *International Journal of Environmental Research and Public Health*, 17(21):7872, October 2020. ISSN 1660-4601. doi: 10.3390/ijerph17217872.

[264] Wesley L. Passos, Gabriel M. Araujo, Amaro A. De Lima, Sergio L. Netto, and Eduardo A.B. Da Silva. Automatic detection of Aedes aegypti breeding grounds based on deep networks with spatio-temporal consistency. *Computers,*

Environment and Urban Systems, 93:101754, April 2022. ISSN 01989715. doi: 10.1016/j.compenvurbsys.2021.101754. URL https://linkinghub.elsevier.com/retrieve/pii/S0198971521001617.

[265] Moritz U G Kraemer, Nuno R Faria, Robert C Reiner, Nick Golding, Birgit Nikolay, Stephanie Stasse, Michael A Johansson, Henrik Salje, Ousmane Faye, G R William Wint, Matthias Niedrig, Freya M Shearer, Sarah C Hill, Robin N Thompson, Donal Bisanzio, Nuno Taveira, Heinrich H Nax, Bary S R Pradelski, Elaine O Nsoesie, Nicholas R Murphy, Isaac I Bogoch, Kamran Khan, John S Brownstein, Andrew J Tatem, Tulio De Oliveira, David L Smith, Amadou A Sall, Oliver G Pybus, Simon I Hay, and Simon Cauchemez. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. *The Lancet Infectious Diseases*, 17(3):330–338, March 2017. ISSN 14733099. doi: 10.1016/S1473-3099(16)30513-8. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309916305138.

[266] Natalya Kostandova, Ronan Corgel, Shweta Bansal, Sophie Bérubé, Eimear Cleary, Chelsea Hansen, Matt D. T. Hitchings, Bernardo García-Carreras, Lauren Gardner, Moritz U. G. Kraemer, Shengjie Lai, Yao Li, Amanda C. Perofsky, Giulia Pullano, Jonathan M. Read, Gabriel Ribeiro Dos Santos, Henrik Salje, Saki Takahashi, Cécile Viboud, Jasmine Wang, Derek A. T. Cummings, and Amy Wesolowski. Improving mobility data for infectious disease research. *Nature Human Behaviour*, March 2025. ISSN 2397-3374. doi: 10.1038/s41562-025-02151-3. URL https://www.nature.com/articles/s41562-025-02151-3.

[267] Natalya Kostandova, Christine Prosperi, Simon Mutembo, Chola Nakazwe, Harriet Namukoko, Bertha Nachinga, Shengjie Lai, Andrew J. Tatem, Qianwen Duan, Elliot N. Kabalo, Kabondo Makungo, Gershom Chongwe, Innocent Chilumba, Gloria Musukwa, Kalumbu H. Matakala, Mutinta Hamahuwa, Webster Mufwambi, Japhet Matoba, Irene Mutale, Kenny Situtu, Edgar Simulundu, Phillimon Ndubani, Alvira Z. Hasan, Shaun A. Truelove, Amy K. Winter, Andrea C. Carcelen, Bryan Lau, William J. Moss, and Amy Wesolowski. Comparing and integrating human mobility data sources for measles transmission modeling in Zambia. *PLOS Global Public Health*, 5(5):e0003906, May 2025. ISSN 2767-3375. doi: 10.1371/journal.pgph.0003906. URL https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0003906. Publisher: Public Library of Science.

[268] Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore Y Piontti, Natalie E. Dean, Diana Patricia Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, Margaret Bray, M. Elizabeth Halloran, Ira M. Longini, and Alessandro Vespignani. Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences*, 114(22), May 2017. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1620161114. URL https://pnas.org/doi/full/10.1073/pnas.1620161114.

[269] Pierre Nouvellet, Sangeeta Bhatia, Anne Cori, Kylie E. C. Ainslie, Marc Baguelin, Samir Bhatt, Adhiratha Boonyasiri, Nicholas F. Brazeau, Lorenzo Cattarino, Laura V. Cooper, Helen Coupland, Zulma M. Cucunuba, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra A. Djaafara, Ilaria Dorigatti, Oliver D. Eales, Sabine L. Van Elsland, Fabricia F. Nascimento, Richard G. FitzJohn, Katy A. M. Gaythorpe, Lily Geidelberg, William D. Green, Arran Hamlet, Katharina Hauck, Wes Hinsley, Natsuko Imai, Benjamin Jeffrey, Edward Knock, Daniel J. Laydon, John A. Lees, Tara Mangal, Thomas A. Mellan, Gemma Nedjati-Gilani, Kris V. Parag, Margarita Pons-Salort, Manon Ragonnet-Cronin, Steven Riley, H. Juliette T. Unwin, Robert Verity, Michaela A. C. Vollmer, Erik Volz, Patrick G. T. Walker, Caroline E. Walters, Haowei Wang, Oliver J. Watson, Charles Whittaker, Lilith K. Whittles, Xiaoyue Xi, Neil M. Ferguson, and Christl A. Donnelly. Reduction in mobility and COVID-19 transmission. *Nature Communications*, 12(1):1090, February 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21358-2. URL https://www.nature.com/articles/s41467-021-21358-2.

[270] Google. Community Mobility Reports, October 2022. URL https://www.google.com/covid19/mobility/.

[271] Apple. Mobility Trends Reports, 2022. URL https://github.com/rearc-data/apple-maps-mobility-trends-covid-19.

[272] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, May 2020. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abb6936. URL https://www.science.org/doi/10.1126/science.abb6936.

[273] Ilaria Dorigatti, Arran Hamlet, Ricardo Aguas, Lorenzo Cattarino, Anne Cori, Christl A Donnelly, Tini Garske, Natsuko Imai, and Neil M Ferguson. International risk of yellow fever spread from the ongoing outbreak in Brazil, December 2016 to May 2017. *Eurosurveillance*, 22(28):30572, July 2017. ISSN 1025-496X. doi: 10.2807/1560-7917.ES.2017.22.28.30572. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5545764/.

[274] Caroline O. Buckee, Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, Yonatan H. Grad, Bryan Grenfell, M. Elizabeth Halloran, Moritz U. G. Kraemer, Marc Lipsitch, C. Jessica E. Metcalf, Lauren Ancel Meyers, T. Alex Perkins, Mauricio Santillana, Samuel V. Scarpino, Cecile Viboud, Amy Wesolowski, and Andrew Schroeder. Aggregated mobility data could help fight COVID-19. *Science*, 368(6487):145–146, April 2020. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abb8021. URL https://www.science.org/doi/10.1126/science.abb8021.

[275] Amy Wesolowski, Caroline O. Buckee, Kenth Engø-Monsen, and C. J. E. Metcalf. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. *Journal of Infectious Diseases*, 214(suppl 4):S414–S420, December 2016. ISSN 0022-1899, 1537-6613. doi: 10.1093/infdis/jiw273. URL https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiw273.

[276] Bouke C. De Jong, Badou M. Gaye, Jeroen Luyten, Bart Van Buitenen, Emmanuel André, Conor J. Meehan, Cian O'Siochain, Kristyna Tomsu, Jérôme Urbain, Koen Peeters Grietens, Maureen Njue, Wim Pinxten, Florian Gehre, Ousman Nyan, Anne Buvé, Anna Roca, Raffaella Ravinetto, and Martin Antonio. Ethical Considerations for Movement Mapping to Identify Disease Transmission Hotspots. *Emerging Infectious Diseases*, 25(7), July 2019. ISSN 1080-6040, 1080-6059. doi: 10.3201/eid2507.181421. URL http://wwwnc.cdc.gov/eid/article/25/7/18-1421_article.htm.

[277] Lydia H V Franklinos, Kate E Jones, David W Redding, and Ibrahim Abubakar. The effect of global change on mosquito-borne disease. *The Lancet Infectious Diseases*, 19(9):e302–e312, September 2019. ISSN 14733099. doi: 10.1016/S1473-3099(19)30161-6. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309919301616.

[278] Shovanur Haque, Kerrie Mengersen, Ian Barr, Liping Wang, Weizhong Yang, Sotiris Vardoulakis, Hilary Bambrick, and Wenbiao Hu. Towards development of functional climate-driven early warning systems for climate-sensitive infectious diseases: Statistical models and recommendations. *Environmental Research*, 249:118568, May 2024. ISSN 0013-9351. doi: 10.1016/j.envres.2024.118568. URL https://www.sciencedirect.com/science/article/pii/S0013935124004729.

[279] Sadie J Ryan, Catherine A Lippi, Talia Caplan, Avriel Diaz, Willy Dunbar, Shruti Grover, Simon Johnson, Rebecca Knowles, Rachel Lowe, Bilal A Mateen, Madeleine C Thomson, and Anna M Stewart-Ibarra. The current landscape of software tools for the climate-sensitive infectious disease modelling community. *The Lancet Planetary Health*, 7(6):e527–e536, June 2023. ISSN 25425196. doi: 10.1016/S2542-5196(23)00056-6. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519623000566.

[280] Dung Phung, Felipe J. Colón-González, Daniel M. Weinberger, Vinh Bui, Son Nghiem, Cordia Chu, Hai Phung, Nam Sinh Vu, Quang-Van Doan, Masahiro Hashizume, Colleen L. Lau, Simon Reid, Lan Trong Phan, Duong Nhu

Tran, Cong Tuan Pham, Kien Quoc Do, and Robert Dubrow. Advancing adoptability and sustainability of digital prediction tools for climate-sensitive infectious disease prevention and control. *Nature Communications*, 16(1):1644, February 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-56826-6. URL https://www.nature.com/articles/s41467-025-56826-6.

[281] Stephen E. Fick and Robert J. Hijmans. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315, October 2017. ISSN 0899-8418, 1097-0088. doi: 10.1002/joc.5086. URL https://onlinelibrary.wiley.com/doi/10.1002/joc.5086.

[282] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia De Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730): 1999–2049, July 2020. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.3803. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3803.

[283] Thomas Reichler and Junsu Kim. Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model. *Journal of Geophysical Research: Atmospheres*, 113(D5):2007JD009278, March 2008. ISSN 0148-0227. doi: 10.1029/2007JD009278. URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2007JD009278.

[284] Kingsley Eghonghon Ukhurebor, Samuel Ogochukwu Azi, Uyiosa Osagie Aigbe, Robert Birundu Onyancha, and Joseph Onyeka Emegha. Analyzing the uncertainties between reanalysis meteorological data and ground measured meteorological data. *Measurement*, 165:108110, December 2020. ISSN 02632241. doi: 10.1016/j.measurement.2020.108110. URL https://linkinghub.elsevier.com/retrieve/pii/S0263224120306485.

[285] Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, December 2023. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adi2336. URL https://www.science.org/doi/10.1126/science.adi2336.

[286] Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R. Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, Remi Lam, and Matthew Willson. Probabilistic weather forecasting with machine learning. *Nature*, December 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-08252-9. URL https://www.nature.com/articles/s41586-024-08252-9.

[287] K. Kashinath, M. Mustafa, A. Albert, J-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and Prabhat. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200093, April 2021. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2020.0093. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0093.

[288] Kelly R. Moran, Geoffrey Fairchild, Nicholas Generous, Kyle Hickmann, Dave Osthus, Reid Priedhorsky, James Hyman, and Sara Y. Del Valle. Epidemic Forecasting is Messier Than Weather Forecasting: The Role of Human Behavior and Internet Data Streams in Epidemic Forecast. *Journal of Infectious Diseases*, 214(suppl 4):S404–S408, December 2016. ISSN 0022-1899, 1537-6613. doi: 10.1093/infdis/jiw375. URL https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiw375.

[289] Yang Ye, Abhishek Pandey, Carolyn Bawden, Dewan Md Sumsuzzman, Rimpi Rajput, Affan Shoukat, Burton H. Singer, Seyed M. Moghadas, and Alison P. Galvani. Integrating artificial intelligence with mechanistic epidemiological modeling: a scoping review of opportunities and challenges. *Nature Communications*, 16(1):581, January 2025. ISSN 2041-1723. doi: 10.1038/s41467-024-55461-x. URL https://www.nature.com/articles/s41467-024-55461-x. Publisher: Nature Publishing Group.

[290] Leo Heijnen and Gertjan Medema. Surveillance of Influenza A and the pandemic influenza A (H1N1) 2009 in sewage and surface water in the Netherlands. *Journal of Water and Health*, 9(3):434–442, September 2011. ISSN 1477-8920, 1996-7829. doi: 10.2166/wh.2011.019. URL https://iwaponline.com/jwh/article/9/3/434/19808/Surveillance-of-Influenza-A-and-the-pandemic.

[291] Radboud J. Duintjer Tebbens, Marita Zimmermann, Mark A. Pallansch, and Kimberly M. Thompson. Insights from a Systematic Search for Information on Designs, Costs, and Effectiveness of Poliovirus Environmental Surveillance Systems. *Food and Environmental Virology*, 9(4):361–382, December 2017. ISSN 1867-0334, 1867-0342. doi: 10.1007/s12560-017-9314-4. URL http://link.springer.com/10.1007/s12560-017-9314-4.

[292] Phil M. Choi, Ben J. Tscharke, Erica Donner, Jake W. O'Brien, Sharon C. Grant, Sarit L. Kaserzon, Rachel Mackie, Elissa O'Malley, Nicholas D. Crosbie, Kevin V. Thomas, and Jochen F. Mueller. Wastewater-based epidemiology biomarkers: Past, present and future. *TrAC Trends in Analytical Chemistry*, 105:453–469, August 2018. ISSN 01659936. doi: 10.1016/j.trac.2018.06.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0165993618301444.

[293] Fuqing Wu, Amy Xiao, Jianbo Zhang, Katya Moniz, Noriko Endo, Federica Armas, Mary Bushman, Peter R. Chai, Claire Duvallet, Timothy B. Erickson, Katelyn Foppe, Newsha Ghaeli, Xiaoqiong Gu, William P. Hanage, Katherine H. Huang, Wei Lin Lee, Kyle A. McElroy, Steven F. Rhode, Mariana Matus, Stefan Wuertz, Janelle Thompson, and Eric J. Alm. Wastewater surveillance of SARS-CoV-2 across 40 U.S. states from February to June 2020. *Water Research*, 202:117400, September 2021. ISSN 00431354. doi: 10.1016/j.watres.2021.117400. URL https://linkinghub.elsevier.com/retrieve/pii/S0043135421005984.

[294] Luke S. Hillary, Kata Farkas, Kathryn H. Maher, Anita Lucaci, Jamie Thorpe, Marco A. Distaso, William H. Gaze, Steve Paterson, Terry Burke, Thomas R. Connor, James E. McDonald, Shelagh K. Malham, and David L. Jones. Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. *Water Research*, 200:117214, July 2021. ISSN 0043-1354. doi: 10.1016/j.watres.2021.117214. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8105641/.

[295] Nicole Acosta, María A. Bautista, Jordan Hollman, Janine McCalder, Alexander Buchner Beaudet, Lawrence Man, Barbara J. Waddell, Jianwei Chen, Carmen Li, Darina Kuzma, Srijak Bhatnagar, Jenine Leal, Jon Meddings, Jia Hu, Jason L. Cabaj, Norma J. Ruecker, Christopher Naugler, Dylan R. Pillai, Gopal Achari, M. Cathryn Ryan, John M. Conly, Kevin Frankowski, Casey RJ Hubert, and Michael D. Parkins. A multicenter study investigating SARS-CoV-2 in tertiary-care hospital wastewater. viral burden correlates with increasing hospitalized cases as well as

hospital-associated transmissions and outbreaks. *Water Research*, 201:117369, August 2021. ISSN 00431354. doi: 10.1016/j.watres.2021.117369. URL https://linkinghub.elsevier.com/retrieve/pii/S0043135421005674.

[296] Warish Ahmed, Aaron Bivins, Stuart L. Simpson, Paul M. Bertsch, John Ehret, Ian Hosegood, Suzanne S. Metcalfe, Wendy J.M. Smith, Kevin V. Thomas, Josh Tynan, and Jochen F. Mueller. Wastewater surveillance demonstrates high predictive value for COVID-19 infection on board repatriation flights to Australia. *Environment International*, 158: 106938, January 2022. ISSN 01604120. doi: 10.1016/j.envint.2021.106938. URL https://linkinghub.elsevier.com/retrieve/pii/S0160412021005638.

[297] Pruthvi Kilaru, Dustin Hill, Kathryn Anderson, Mary B Collins, Hyatt Green, Brittany L Kmush, and David A Larsen. Wastewater Surveillance for Infectious Disease: A Systematic Review. *American Journal of Epidemiology*, 192(2):305–322, February 2023. ISSN 0002-9262. doi: 10.1093/aje/kwac175. URL https://doi.org/10.1093/aje/kwac175.

[298] Sílvia Monteiro, Raquel Pimenta, Filipa Nunes, Mónica V. Cunha, and Ricardo Santos. Detection of dengue virus and chikungunya virus in wastewater in Portugal: an exploratory surveillance study. *The Lancet Microbe*, 5 (11), November 2024. ISSN 2666-5247. doi: 10.1016/S2666-5247(24)00150-2. URL https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(24)00150-2/fulltext. Publisher: Elsevier.

[299] Ocean Thakali, Sunayana Raya, Bikash Malla, Sarmila Tandukar, Ananda Tiwari, Samendra P. Sherchan, Jeevan B. Sherchand, and Eiji Haramoto. Pilot study on wastewater surveillance of dengue virus RNA: Lessons, challenges, and implications for future research. *Environmental Challenges*, 9:100614, December 2022. ISSN 2667-0100. doi: 10.1016/j.envc.2022.100614. URL https://www.sciencedirect.com/science/article/pii/S2667010022001706.

[300] Jessica E. Stockdale, Pengyu Liu, and Caroline Colijn. The potential of genomics for infectious disease forecasting. *Nature Microbiology*, 7(11):1736–1743, October 2022. ISSN 2058-5276. doi: 10.1038/s41564-022-01233-6. URL https://www.nature.com/articles/s41564-022-01233-6.

[301] Verity Hill, Simon Dellicour, Marta Giovanetti, and Nathan D. Grubaugh. Phylogenetic insights into the transmission dynamics of arthropod-borne viruses. *Nature Reviews Genetics*, pages 1–15, June 2025. ISSN 1471-0064. doi: 10.1038/s41576-025-00854-x. URL https://www.nature.com/articles/s41576-025-00854-x. Publisher: Nature Publishing Group.

[302] UKHSA. EMHP wastewater monitoring of SARS-CoV-2 in England: 15 July 2020 to 30 March 2022, May 2022. URL https://www.gov.uk/government/publications/monitoring-of-sars-cov-2-rna-in-england-wastewater-monthly-statistics-15-july-2020-to-30-march-2022/emhp-wastewater-monitoring-of-sars-cov-2-in-england-15-july-2020-to-30-march-2022.

[303] Sarah M. Prasek, Ian L. Pepper, Gabriel K. Innes, Stephanie Slinski, Walter Q. Betancourt, Aidan R. Foster, Hayley D. Yaglom, W. Tanner Porter, David M. Engelthaler, and Bradley W. Schmitz. Variant-specific SARS-CoV-2 shedding rates in wastewater. *Science of The Total Environment*, 857:159165, January 2023. ISSN 00489697. doi: 10.1016/j.scitotenv.2022.159165. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969722062647.

[304] Joseph L.-H. Tsui, John T. McCrone, Ben Lambert, Sumali Bajaj, Rhys P. D. Inward, Paolo Bosetti, Rosario Evans Pena, Houriiyah Tegally, Verity Hill, Alexander E. Zarebski, Thomas P. Peacock, Luyang Liu, Neo Wu, Megan Davis, Isaac I. Bogoch, Kamran Khan, Meaghan Kall, Nurin Iwani Binti Abdul Aziz, Rachel Colquhoun, Áine O'Toole, Ben Jackson, Abhishek Dasgupta, Eduan Wilkinson, Tulio De Oliveira, The COVID-19 Genomics UK (COG-UK) consortium¶, Thomas R. Connor, Nicholas J. Loman, Vittoria Colizza, Christophe Fraser, Erik Volz, Xiang Ji, Bernardo Gutierrez, Meera Chand, Simon Dellicour, Simon Cauchemez, Jayna Raghwani, Marc A. Suchard,

Philippe Lemey, Andrew Rambaut, Oliver G. Pybus, and Moritz U. G. Kraemer. Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1. *Science*, 381(6655):336–343, July 2023. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adg6605. URL https://www.science.org/doi/10.1126/science.adg6605.

[305] Cécile Tran-Kiem and Trevor Bedford. Estimating the reproduction number and transmission heterogeneity from the size distribution of clusters of identical pathogen sequences. *Proceedings of the National Academy of Sciences*, 121 (15):e2305299121, April 2024. doi: 10.1073/pnas.2305299121. URL https://www.pnas.org/doi/10.1073/pnas.2305299121. Publisher: Proceedings of the National Academy of Sciences.

[306] Paul Elliott, David Haw, Haowei Wang, Oliver Eales, Caroline E. Walters, Kylie E. C. Ainslie, Christina Atchison, Claudio Fronterre, Peter J. Diggle, Andrew J. Page, Alexander J. Trotter, Sophie J. Prosolek, The COVID-19 Genomics UK (COG-UK) Consortium, Deborah Ashby, Christl A. Donnelly, Wendy Barclay, Graham Taylor, Graham Cooke, Helen Ward, Ara Darzi, and Steven Riley. REACT-1 round 13 final report: exponential growth, high prevalence of SARS-CoV-2 and vaccine effectiveness associated with Delta variant in England during May to July 2021. preprint, Epidemiology, September 2021. URL http://medrxiv.org/lookup/doi/10.1101/2021.09.02.21262979.

[307] Paul Elliott, Oliver Eales, Nicholas Steyn, David Tang, Barbara Bodinier, Haowei Wang, Joshua Elliott, Matthew Whitaker, Christina Atchison, Peter J. Diggle, Andrew J. Page, Alexander J. Trotter, Deborah Ashby, Wendy Barclay, Graham Taylor, Helen Ward, Ara Darzi, Graham S. Cooke, Christl A. Donnelly, and Marc Chadeau-Hyam. Twin peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England. *Science*, 376(6600):eabq4411, June 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abq4411. URL https://www.science.org/doi/10.1126/science.abq4411.

[308] Paul Elliott, Matthew Whitaker, David Tang, Oliver Eales, Nicholas Steyn, Barbara Bodinier, Haowei Wang, Joshua Elliott, Christina Atchison, Deborah Ashby, Wendy Barclay, Graham Taylor, Ara Darzi, Graham S. Cooke, Helen Ward, Christl A. Donnelly, Steven Riley, and Marc Chadeau-Hyam. Design and Implementation of a National SARS-CoV-2 Monitoring Program in England: REACT-1 Study. *American Journal of Public Health*, 113(5):545–554, May 2023. ISSN 0090-0036, 1541-0048. doi: 10.2105/AJPH.2023.307230. URL https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2023.307230.

[309] Steven Riley, Christina Atchison, Deborah Ashby, Christl A. Donnelly, Wendy Barclay, Graham S. Cooke, Helen Ward, Ara Darzi, Paul Elliott, and REACT study group. REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: Study protocol. *Wellcome Open Research*, 5:200, April 2021. ISSN 2398-502X. doi: 10.12688/wellcomeopenres.16228.2. URL https://wellcomeopenresearch.org/articles/5-200/v2.

[310] C Jessica E Metcalf, Jeremy Farrar, Felicity T Cutts, Nicole E Basta, Andrea L Graham, Justin Lessler, Neil M Ferguson, Donald S Burke, and Bryan T Grenfell. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *The Lancet*, 388(10045):728–730, August 2016. ISSN 01406736. doi: 10.1016/S0140-6736(16)30164-7. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673616301647.

[311] Qinlong Jing, Yilan Li, Jianhua Liu, Liyun Jiang, Zongqiu Chen, Wenzhe Su, Guthrie S Birkhead, Jiahai Lu, and Zhicong Yang. Dengue Underestimation in Guangzhou, China: Evidence of Seroprevalence in Communities With No Reported Cases Before a Large Outbreak in 2014. *Open Forum Infectious Diseases*, 6(7):ofz256, July 2019. ISSN 2328-8957. doi: 10.1093/ofid/ofz256. URL https://doi.org/10.1093/ofid/ofz256.

[312] Akhilesh C. Mishra, Vidya A. Arankalle, Swapnil A. Gadhave, Pritam H. Mahadik, Shubham Shrivastava, Mandar Bhutkar, and Varsha M. Vaidya. Stratified sero-prevalence revealed overall high disease burden of dengue but suboptimal

immunity in younger age groups in Pune, India. *PLoS Neglected Tropical Diseases*, 12(8):e0006657, August 2018. ISSN 1935-2727. doi: 10.1371/journal.pntd.0006657. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6095695/.

[313] Angkana T. Huang, Darunee Buddhari, Surachai Kaewhiran, Sopon Iamsirithaworn, Direk Khampaen, Aaron Farmer, Stefan Fernandez, Stephen J. Thomas, Isabel Rodriguez-Barraquer, Taweewun Hunsawong, Anon Srikiatkhachorn, Gabriel Ribeiro Dos Santos, Megan O'Driscoll, Marco Hamins-Puertolas, Timothy Endy, Alan L. Rothman, Derek A. T. Cummings, Kathryn Anderson, and Henrik Salje. Reconciling heterogeneous dengue virus infection risk estimates from different study designs. *Proceedings of the National Academy of Sciences*, 122(1):e2411768121, January 2025. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2411768121. URL https://pnas.org/doi/10.1073/pnas.2411768121.

[314] Everlyn Kamau, Junjie Chen, Sumali Bajaj, Nicolás Torres, Richard Creswell, Jaime A. Pavlich-Mariscal, Christl Donnelly, Zulma Cucunubá, and Ben Lambert. The Mathematics of Serocatalytic Models with applications to public health data, January 2025. URL http://medrxiv.org/lookup/doi/10.1101/2025.01.21.25320864.

[315] Anna Vicco, Clare McCormack, Belen Pedrique, Isabela Ribeiro, Gathsaurie Neelika Malavige, and Ilaria Dorigatti. A scoping literature review of global dengue age-stratified seroprevalence data: estimating dengue force of infection in endemic countries. *eBioMedicine*, 104:105134, June 2024. ISSN 23523964. doi: 10.1016/j.ebiom.2024.105134. URL https://linkinghub.elsevier.com/retrieve/pii/S2352396424001695.

[316] Mairead G. Whelan, Harriet Ware, Himanshu Ranka, Sean Kenny, Sabah Shaikh, Yannik Roell, Shaila Akter, Anabel Selemon, Emilie Toews, May Chu, Niklas Bobrovitz, Rahul K. Arora, and Thomas Jaenisch. ArboTracker: a multipathogen dashboard and data platform for arbovirus seroprevalence studies. *The Lancet. Infectious Diseases*, 24 (11):e670–e671, November 2024. ISSN 1474-4457. doi: 10.1016/S1473-3099(24)00585-1.

[317] Rahul K Arora, Abel Joseph, Jordan Van Wyk, Simona Rocco, Austin Atmaja, Ewan May, Tingting Yan, Niklas Bobrovitz, Jonathan Chevrier, Matthew P Cheng, Tyler Williamson, and David L Buckeridge. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *The Lancet Infectious Diseases*, 21(4):e75–e76, April 2021. ISSN 14733099. doi: 10.1016/S1473-3099(20)30631-9. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309920306319.

[318] Zulma M. Cucunubá, Nicolás T. Domínguez, Ben Lambert, and Pierre Nouvellet. *serofoi: Bayesian Estimation of the Force of Infection from Serological Data*. 2025. URL https://github.com/epiverse-trace/serofoi.

[319] Nathanaël Hozé, Margarita Pons-Salort, C. Jessica E. Metcalf, Michael White, Henrik Salje, and Simon Cauchemez. RSero: A user-friendly R package to reconstruct pathogen circulation history from seroprevalence studies. *PLOS Computational Biology*, 21(2):e1012777, February 2025. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012777. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012777. Publisher: Public Library of Science.

[320] Olivia Man, Alicia Kraay, Ruth Thomas, James Trostle, Gwenyth O. Lee, Charlotte Robbins, Amy C. Morrison, Josefina Coloma, and Joseph N. S. Eisenberg. Characterizing dengue transmission in rural areas: A systematic review. *PLOS Neglected Tropical Diseases*, 17(6):e0011333, June 2023. ISSN 1935-2735. doi: 10.1371/journal.pntd.0011333. URL https://dx.plos.org/10.1371/journal.pntd.0011333.

[321] Zhiyuan Chen, Philippe Lemey, and Hongjie Yu. Approaches and challenges to inferring the geographical source of infectious disease outbreaks using genomic data. *The Lancet Microbe*, 5(1):e81–e92, January 2024. ISSN 2666-5247. doi: 10.1016/S2666-5247(23)00296-3. URL https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(23)00296-3/fulltext. Publisher: Elsevier.

[322] Oliver J. Brady, Leonardo S. Bastos, Jamie M. Caldwell, Simon Cauchemez, Hannah E. Clapham, Illaria Dorigatti, Katy A. M. Gaythorpe, Wenbiao Hu, Laith Hussain-Alkhateeb, Michael A. Johansson, Ahyoung Lim, Velma K. Lopez, Richard James Maude, Jane P. Messina, Erin A. Mordecai, Andrew Townsend Peterson, Isabel Rodriquez-Barraquer, Ingrid B. Rabe, Diana P. Rojas, Sadie J. Ryan, Henrik Salje, Jan C. Semenza, and Quan Minh Tran. Why the growth of arboviral diseases necessitates a new generation of global risk maps and future projections. *PLOS Computational Biology*, 21(4):e1012771, April 2025. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012771. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012771. Publisher: Public Library of Science.

[323] Daniel Bernoulli. Réflexions sur les avantages de l'inoculation. Lue dans l'assemblée publique du 16.4.1760 : Extrait du Mercure de France. 1760. doi: 10.7891/E-MANUSCRIPTA-148857. URL https://www.e-manuscripta.ch/doi/10.7891/e-manuscripta-148857. Medium: 8 S. ; 24 x 18,5 cm Publisher: [s.n.].

[324] Francis Eg Cox. History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors*, 3(1):5, February 2010. ISSN 1756-3305. doi: 10.1186/1756-3305-3-5.

[325] Ronald Ross. *The Prevention of malaria*. New York, E.P. Dutton & company, 1910. URL https://books.google.com/books?id=0jRaWNX--s0C.

[326] Patrick Manson. On the Development of Filaria sanguinis hominis, and on the Mosquito considered as a Nurse*. *Journal of the Linnean Society of London, Zoology*, 14(75):304–311, 1878. ISSN 1096-3642. doi: 10.1111/j.1096-3642.1878.tb01837.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1096-3642.1878.tb01837.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1096-3642.1878.tb01837.x.

[327] Ronald Ross. An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part I. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 92(638):204–230, 1916. ISSN 09501207. URL http://www.jstor.org/stable/93760. Publisher: The Royal Society.

[328] Ronald Ross and Hilda P. Hudson. An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part II. *Proceedings of the Royal Society of London Series A*, 93:212–225, May 1917. ISSN 0080-46301364-5021. doi: 10.1098/rspa.1917.0014. URL https://ui.adsabs.harvard.edu/abs/1917RSPSA..93..212R. ADS Bibcode: 1917RSPSA..93..212R.

[329] Ronald Ross and Hilda P. Hudson. An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part III. *Proceedings of the Royal Society of London Series A*, 93:225–240, May 1917. ISSN 0080-46301364-5021. doi: 10.1098/rspa.1917.0015. URL https://ui.adsabs.harvard.edu/abs/1917RSPSA..93..225R. ADS Bibcode: 1917RSPSA..93..225R.

[330] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115(772):700–721, 1927. Publisher: The Royal Society London.

[331] John Snow. *On the Mode of Communication of Cholera*. John Churchill, 1855.

[332] William Farr. Causes of death in England and Wales. In *Second Annual Report of the Registrar General of Births, Deaths and Marriages in England.*, pages 69–98. 1840.

[333] Justin Lessler and Derek A. T. Cummings. Mechanistic Models of Infectious Disease and Their Impact on Public Health. *American Journal of Epidemiology*, 183(5):415–422, March 2016. ISSN 0002-9262. doi: 10.1093/aje/kww021. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5006438/.

[334] Tal T. Robin, Jaime Cascante-Vega, Jeffrey Shaman, and Sen Pei. System identifiability in a time-evolving agent-based model. *PLOS ONE*, 19(1):e0290821, January 2024. ISSN 1932-6203. doi: 10.1371/journal.pone.0290821. URL https://dx.plos.org/10.1371/journal.pone.0290821.

[335] Katia Koelle, Alex P Farrell, Christopher B Brooke, and Ruian Ke. Within-host infectious disease models accommodating cellular coinfection, with an application to influenza. *Virus Evolution*, 5(2):vez018, July 2019. ISSN 2057-1577. doi: 10.1093/ve/vez018. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6613536/.

[336] Maíra Aguiar, Vizda Anam, Konstantin B. Blyuss, Carlo Delfin S. Estadilla, Bruno V. Guerrero, Damián Knopoff, Bob W. Kooi, Akhil Kumar Srivastav, Vanessa Steindorf, and Nico Stollenwerk. Mathematical models for dengue fever epidemiology: A 10-year systematic review. *Physics of Life Reviews*, 40:65–92, March 2022. ISSN 15710645. doi: 10.1016/j.plrev.2022.02.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1571064522000045.

[337] Kathleen K. Steinhöfel, David Heslop, and C. Raina MacIntyre. Agent-Based Models of Virus Infection. *Current Clinical Microbiology Reports*, 12(1):2, January 2025. ISSN 2196-5471. doi: 10.1007/s40588-024-00238-5. URL https://doi.org/10.1007/s40588-024-00238-5.

[338] Matt J Keeling and Ken T.D Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4): 295–307, September 2005. ISSN 1742-5689. doi: 10.1098/rsif.2005.0051. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1578276/.

[339] Donald S. Burke, Joshua M. Epstein, Derek A. T. Cummings, Jon I. Parker, Kenneth C. Cline, Ramesh M. Singa, and Shubha Chakravarty. Individual-based computational modeling of smallpox epidemic control strategies. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, 13(11):1142–1149, November 2006. ISSN 1553-2712. doi: 10.1197/j.aem.2006.07.017.

[340] C. C. Graciani Rodrigues, Aquino L. Espíndola, and T. J. P. Penna. An agent-based computational model for tuberculosis spreading on age-structured populations. *Physica A: Statistical Mechanics and its Applications*, 428: 52–59, June 2015. ISSN 0378-4371. doi: 10.1016/j.physa.2015.02.027. URL https://www.sciencedirect.com/science/article/pii/S0378437115001296.

[341] Dhananjai M. Rao, Alexander Chernyakhovsky, and Victoria Rao. Modeling and analysis of global epidemiology of avian influenza. *Environmental Modelling & Software*, 24(1):124–134, January 2009. ISSN 1364-8152. doi: 10.1016/j.envsoft.2008.06.011. URL https://www.sciencedirect.com/science/article/pii/S1364815208001072.

[342] Robert Hinch, Jasmina Panovska-Griffiths, William J. M. Probert, Luca Ferretti, Chris Wymant, Francesco Di Lauro, Nikolas Baya, Mahan Ghafari, Lucie Abeler-Dörner, The COVID-19 Genomics UK (COG-UK) Consortium, and Christophe Fraser. Estimating SARS-CoV-2 variant fitness and the impact of interventions in England using statistical and geo-spatial agent-based models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2233):20210304, October 2022. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2021.0304. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2021.0304.

[343] Robert Hinch, William J. M. Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferretti, Daniel Montero, James Warren, Nicole Mather, Matthew Abueg, Neo Wu, Olivier Legat, Katie Bentley, Thomas Mead, Kelvin Van-Vuuren, Dylan Feldner-Busztin, Tommaso Ristori, Anthony Finkelstein, David G. Bonsall, Lucie Abeler-Dörner, and Christophe Fraser. OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *PLOS Computational Biology*, 17(7):e1009146, July 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009146. URL https://dx.plos.org/10.1371/journal.pcbi.1009146.

[344] Olga Krivorotko, Mariia Sosnovskaia, Ivan Vashchenko, Cliff Kerr, and Daniel Lesnic. Agent-based modeling of COVID-19 outbreaks for New York state and UK: Parameter identification algorithm. *Infectious Disease Modelling*, 7(1):30–44, March 2022. ISSN 24680427. doi: 10.1016/j.idm.2021.11.004. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042721000798.

[345] Takashi Shiono. Estimation of agent-based models using Bayesian deep learning approach of BayesFlow. *Journal of Economic Dynamics and Control*, 125:104082, April 2021. ISSN 01651889. doi: 10.1016/j.jedc.2021.104082. URL https://linkinghub.elsevier.com/retrieve/pii/S0165188921000178.

[346] Josie McCulloch, Jiaqi Ge, Jonathan A. Ward, Alison Heppenstall, J. Gareth Polhill, and Nick Malleson. Calibrating Agent-Based Models Using Uncertainty Quantification Methods. *Journal of Artificial Societies and Social Simulation*, 25(2):1, 2022. ISSN 1460-7425. doi: 10.18564/jasss.4791. URL https://www.jasss.org/25/2/1.html.

[347] Nikita Sivakumar, Cameron Mura, and Shayn M. Peirce. Innovations in integrating machine learning and agent-based modeling of biomedical systems. *Frontiers in Systems Biology*, 2:959665, November 2022. ISSN 2674-0702. doi: 10.3389/fsysb.2022.959665. URL https://www.frontiersin.org/articles/10.3389/fsysb.2022.959665/full.

[348] Ruth E. Baker, Jose-Maria Peña, Jayaratnam Jayamohan, and Antoine Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, 14(5):20170660, May 2018. ISSN 1744-9561, 1744-957X. doi: 10.1098/rsbl.2017.0660. URL https://royalsocietypublishing.org/doi/10.1098/rsbl.2017.0660.

[349] Linda J.S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2):128–142, May 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.03.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042716300495.

[350] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, Nicholas Davies, Amy Gimma, Kevin Van Zandvoort, Hamish Gibbs, Joel Hellewell, Christopher I Jarvis, Sam Clifford, Billy J Quilty, Nikos I Bosse, Sam Abbott, Petra Klepac, and Stefan Flasche. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(5):553–558, May 2020. ISSN 14733099. doi: 10.1016/S1473-3099(20)30144-4. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309920301444.

[351] Christovam Barcellos, Vanderlei Matos, Raquel Martins Lana, and Rachel Lowe. Climate change, thermal anomalies, and the recent progression of dengue in Brazil. *Scientific Reports*, 14(1):5948, March 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-56044-y. URL https://www.nature.com/articles/s41598-024-56044-y. Publisher: Nature Publishing Group.

[352] Manisha A. Kulkarni, Claudia Duguay, and Katarina Ost. Charting the evidence for climate change impacts on the global spread of malaria and dengue and adaptive responses: a scoping review of reviews. *Globalization and Health*, 18(1):1, December 2022. ISSN 1744-8603. doi: 10.1186/s12992-021-00793-2. URL https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-021-00793-2.

[353] Yawen Wang, Yuchen Wei, Kehang Li, Xiaoting Jiang, Conglu Li, Qianying Yue, Benny Chung-ying Zee, and Ka Chun Chong. Impact of extreme weather on dengue fever infection in four Asian countries: A modelling analysis. *Environment International*, 169:107518, November 2022. ISSN 01604120. doi: 10.1016/j.envint.2022.107518. URL https://linkinghub.elsevier.com/retrieve/pii/S0160412022004457.

[354] Xiao Wei Sylvia Gwee, Pearleen Ee Yong Chua, and Junxiong Pang. Global dengue importation: a systematic review. *BMC Infectious Diseases*, 21:1078, October 2021. ISSN 1471-2334. doi: 10.1186/s12879-021-06740-1. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8524397/.

[355] Claudia T. Codeco, Sara S. Oliveira, Danielle A.C. Ferreira, Thais I.S. Riback, Leonardo S. Bastos, Raquel M. Lana, Iasmim F. Almeida, Vinicius B. Godinho, Oswaldo G. Cruz, and Flavio C. Coelho. Fast expansion of dengue in Brazil. *The Lancet Regional Health - Americas*, 12:100274, August 2022. ISSN 2667193X. doi: 10.1016/j.lana.2022.100274. URL https://linkinghub.elsevier.com/retrieve/pii/S2667193X22000916.

[356] Christovam Barcellos and Rachel Lowe. Expansion of the dengue transmission area in Brazil: the role of climate and cities. *Tropical medicine & international health: TM & IH*, 19(2):159–168, February 2014. ISSN 1365-3156. doi: 10.1111/tmi.12227.

[357] Clara Champagne and Bernard Cazelles. Comparison of stochastic and deterministic frameworks in dengue modelling. *Mathematical Biosciences*, 310:1–12, April 2019. ISSN 0025-5564. doi: 10.1016/j.mbs.2019.01.010. URL https://www.sciencedirect.com/science/article/pii/S0025556418301652.

[358] Richard Levins. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control1. *Bulletin of the Entomological Society of America*, 15(3):237–240, September 1969. ISSN 0013-8754. doi: 10.1093/besa/15.3.237. URL https://doi.org/10.1093/besa/15.3.237.

[359] James Truscott and Neil M. Ferguson. Evaluating the Adequacy of Gravity Models as a Description of Human Mobility for Epidemic Modelling. *PLoS Computational Biology*, 8(10):e1002699, October 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002699. URL https://dx.plos.org/10.1371/journal.pcbi.1002699.

[360] István Z. Kiss, Joel C. Miller, and Péter L. Simon. *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, volume 46 of *Interdisciplinary Applied Mathematics*. Springer International Publishing, Cham, 2017. ISBN 978-3-319-50804-7 978-3-319-50806-1. doi: 10.1007/978-3-319-50806-1. URL http://link.springer.com/10.1007/978-3-319-50806-1.

[361] Ross-William S. Hendron and Michael B. Bonsall. The interplay of vaccination and vector control on small dengue networks. *Journal of Theoretical Biology*, 407:349–361, October 2016. ISSN 0022-5193. doi: 10.1016/j.jtbi.2016.07.034. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5016021/.

[362] Sebastian Funk, Anton Camacho, Adam J. Kucharski, Rosalind M. Eggo, and W. John Edmunds. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics*, 22:56–61, March 2018. ISSN 1755-4365. doi: 10.1016/j.epidem.2016.11.003. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5871642/.

[363] Max S. Y. Lau, Alex Becker, Wyatt Madden, Lance A. Waller, C. Jessica E. Metcalf, and Bryan T. Grenfell. Comparing and linking machine learning and semi-mechanistic models for the predictability of endemic measles dynamics. *PLOS Computational Biology*, 18(9):e1010251, September 2022. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1010251. URL https://dx.plos.org/10.1371/journal.pcbi.1010251.

[364] Samir Bhatt, Neil Ferguson, Seth Flaxman, Axel Gandy, Swapnil Mishra, and James A. Scott. Semi-Mechanistic Bayesian Modeling of COVID-19 with Renewal Processes, December 2020. URL http://arxiv.org/abs/2012.00394. arXiv:2012.00394 [stat].

[365] James A. Scott, Axel Gandy, Swapnil Mishra, Samir Bhatt, Seth Flaxman, H. Juliette T. Unwin, and Jonathan Ish-Horowicz. Epidemia: An R Package for Semi-Mechanistic Bayesian Modelling of Infectious Diseases using Point Processes, October 2021. URL http://arxiv.org/abs/2110.12461. arXiv:2110.12461 [stat].

[366] Adrian Lison, Sam Abbott, Jana Huisman, and Tanja Stadler. Generative Bayesian modeling to nowcast the effective reproduction number from line list data with missing symptom onset dates. *PLOS Computational Biology*, 20(4): e1012021, April 2024. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012021. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012021. Publisher: Public Library of Science.

[367] X. Sheldon Lin and Gordon E. Willmot. Analysis of a defective renewal equation arising in ruin theory. *Insurance: Mathematics and Economics*, 25(1):63–84, September 1999. ISSN 0167-6687. doi: 10.1016/S0167-6687(99)00026-8. URL https://www.sciencedirect.com/science/article/pii/S0167668799000268.

[368] A. G. McKendrick. Applications of mathematics to Medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, February 1925. ISSN 0013-0915, 1464-3839. doi: 10.1017/S0013091500034428. URL https://www.cambridge.org/core/product/identifier/S0013091500034428/type/journal_article.

[369] William D. Green, Neil M. Ferguson, and Anne Cori. Inferring the reproduction number using the renewal equation in heterogeneous epidemics. *Journal of The Royal Society Interface*, 19(188):20210429, March 2022. ISSN 1742-5662. doi: 10.1098/rsif.2021.0429. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2021.0429.

[370] Ioana Bouros, Robin Thompson, David Gavaghan, and Ben Lambert. The time-dependent reproduction number for epidemics in heterogeneous populations, January 2025. URL http://arxiv.org/abs/2501.17355. arXiv:2501.17355 [q-bio].

[371] Michael George Roberts and Hiroshi Nishiura. Early Estimation of the Reproduction Number in the Presence of Imported Cases: Pandemic Influenza H1N1-2009 in New Zealand. *PLoS ONE*, 6(5):e17835, May 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0017835. URL https://dx.plos.org/10.1371/journal.pone.0017835.

[372] Jacob Curran-Sebastian, Frederik Mølkjær Andersen, and Samir Bhatt. Modelling the stochastic importation dynamics and establishment of novel pathogenic strains using a general branching processes framework. *Mathematical Biosciences*, 380:109352, February 2025. ISSN 0025-5564. doi: 10.1016/j.mbs.2024.109352. URL https://www.sciencedirect.com/science/article/pii/S0025556424002128.

[373] David Champredon, Irena Papst, and Warsame Yusuf. ern: An R package to estimate the effective reproduction number using clinical and wastewater surveillance data. *PLOS ONE*, 19(6):e0305550, June 2024. ISSN 1932-6203. doi: 10.1371/journal.pone.0305550. URL https://dx.plos.org/10.1371/journal.pone.0305550.

[374] B.L. Keyfitz and N. Keyfitz. The McKendrick partial differential equation and its uses in epidemiology and population study. *Mathematical and Computer Modelling*, 26(6):1–9, September 1997. ISSN 08957177. doi: 10.1016/S0895-7177(97)00165-9. URL https://linkinghub.elsevier.com/retrieve/pii/S0895717797001659.

[375] Claudia T. Codeço, Daniel A.M. Villela, and Flavio C. Coelho. Estimating the effective reproduction number of dengue considering temperature-dependent generation intervals. *Epidemics*, 25:101–111, December 2018. ISSN 17554365. doi: 10.1016/j.epidem.2018.05.011. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436517300907.

[376] Jérémie Scire, Jana S. Huisman, Ana Grosu, Daniel C. Angst, Adrian Lison, Jinzhou Li, Marloes H. Maathuis, Sebastian Bonhoeffer, and Tanja Stadler. estimateR: an R package to estimate and monitor the effective reproductive

number. *BMC Bioinformatics*, 24(1):310, August 2023. ISSN 1471-2105. doi: 10.1186/s12859-023-05428-4. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-023-05428-4.

[377] Janet Ong, Stacy Soh, Soon Hoe Ho, Annabel Seah, Borame Sue Dickens, Ken Wei Tan, Joel Ruihan Koo, Alex R. Cook, Daniel R. Richards, Leon Yan-Feng Gaw, Lee Ching Ng, and Jue Tao Lim. Fine-scale estimation of effective reproduction numbers for dengue surveillance. *PLOS Computational Biology*, 18(1):e1009791, January 2022. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009791. URL https://dx.plos.org/10.1371/journal.pcbi.1009791.

[378] Bahman Moraffah. Bayesian Nonparametrics: An Alternative to Deep Learning, March 2024. URL http://arxiv.org/abs/2404.00085. arXiv:2404.00085 [cs, stat].

[379] Richard Creswell, Martin Robinson, David Gavaghan, Kris V. Parag, Chon Lok Lei, and Ben Lambert. A Bayesian nonparametric method for detecting rapid changes in disease transmission. *Journal of Theoretical Biology*, 558: 111351, February 2023. ISSN 00225193. doi: 10.1016/j.jtbi.2022.111351. URL https://linkinghub.elsevier.com/retrieve/pii/S0022519322003423.

[380] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, November 2005. ISBN 978-0-262-25683-4. doi: 10.7551/mitpress/3206.001.0001. URL https://direct.mit.edu/books/book/2320/Gaussian-Processes-for-Machine-Learning.

[381] Gabriel Riutort-Mayol, Paul-Christian Bürkner, Michael R. Andersen, Arno Solin, and Aki Vehtari. Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. *Statistics and Computing*, 33 (1):17, December 2022. ISSN 1573-1375. doi: 10.1007/s11222-022-10167-2. URL https://doi.org/10.1007/s11222-022-10167-2.

[382] Laura M Guzmán-Rincón, Edward M Hill, Louise Dyson, Michael J Tildesley, and Matt J Keeling. Bayesian estimation of real-time epidemic growth rates using Gaussian processes: local dynamics of SARS-CoV-2 in England. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(5):1413–1434, December 2023. ISSN 0035-9254, 1467-9876. doi: 10.1093/jrsssc/qlad056. URL https://academic.oup.com/jrsssc/article/72/5/1413/7209012.

[383] A. Bucci, L. Ippoliti, P. Valentini, and S. Fontanella. Clustering spatio-temporal series of confirmed COVID-19 deaths in Europe. *Spatial Statistics*, 49:100543, June 2022. ISSN 22116753. doi: 10.1016/j.spasta.2021.100543. URL https://linkinghub.elsevier.com/retrieve/pii/S2211675321000531.

[384] Elizaveta Semenova, Yidan Xu, Adam Howes, Theo Rashid, Samir Bhatt, Swapnil Mishra, and Seth Flaxman. PriorVAE: encoding spatial priors with variational autoencoders for small-area estimation. *Journal of The Royal Society Interface*, 19(191):20220094, June 2022. ISSN 1742-5662. doi: 10.1098/rsif.2022.0094. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2022.0094.

[385] Iwona Hawryluk, Henrique Hoeltgebaum, Swapnil Mishra, Xenia Miscouridou, Ricardo P Schnekenberg, Charles Whittaker, Michaela Vollmer, Seth Flaxman, Samir Bhatt, and Thomas A Mellan. Gaussian Process Nowcasting: Application to COVID-19 Mortality Reporting, 2021. URL https://arxiv.org/abs/2102.11249. Version Number: 2.

[386] Shwet Ketu and Pramod Kumar Mishra. Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection. *Applied Intelligence*, 51(3):1492–1512, March 2021. ISSN 1573-7497. doi: 10.1007/s10489-020-01889-9. URL https://doi.org/10.1007/s10489-020-01889-9.

[387] Andrew Gelman and Jennifer Hill. Data Analysis Using Regression and Multilevel/Hierarchical Models, December 2006. URL https://www.cambridge.org/highereducation/books/data-analysis-using-regression-and-multilevel-hierarchical-models/32A29531C7FD730C3A68951A17C9D983. ISBN: 9780511790942 Publisher: Cambridge University Press.

[388] Hanna M. McCaslin, Abigail B. Feuka, and Mevin B. Hooten. Hierarchical computing for hierarchical models in ecology. *Methods in Ecology and Evolution*, 12(2):245–254, February 2021. ISSN 2041-210X, 2041-210X. doi: 10.1111/2041-210X.13513. URL https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13513.

[389] Christopher K. Wikle. Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes. *Ecology*, 84(6):1382–1394, June 2003. ISSN 0012-9658. doi: 10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2. URL http://doi.wiley.com/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2.

[390] James Hensman, Neil D. Lawrence, and Magnus Rattray. Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. *BMC Bioinformatics*, 14(1):252, August 2013. ISSN 1471-2105. doi: 10.1186/1471-2105-14-252. URL https://doi.org/10.1186/1471-2105-14-252.

[391] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Texts in statistical science series. CRC Press, Taylor and Francis Group, Boca Raton London New York, third edition, 2014. ISBN 978-1-4398-4095-5.

[392] Douglas M Watts, Donald S Burke, Bruce A Harrison, Richard E Whitmire, Ananda Nisalak, and others. Effect of temperature on the vector efficiency of Aedes aegypti for dengue 2 virus. *Am J Trop Med Hyg*, 36(1):143–52, 1987. Publisher: Citeseer.

[393] Anna M. Stewart Ibarra, Sadie J. Ryan, Efrain Beltrán, Raúl Mejía, Mercy Silva, and Angel Munoz. Dengue Vector Dynamics (Aedes aegypti) Influenced by Climate and Social Factors in Ecuador: Implications for Targeted Control. *PLoS ONE*, 8(11):e78263, November 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0078263. URL https://dx.plos.org/10.1371/journal.pone.0078263.

[394] A. Gasparrini, B. Armstrong, and M. G. Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 29 (21):2224–2234, September 2010. ISSN 02776715. doi: 10.1002/sim.3940. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.3940.

[395] Latchezar Tomov, Lyubomir Chervenkov, Dimitrina Georgieva Miteva, Hristiana Batselova, and Tsvetelina Velikova. Applications of time series analysis in epidemiology: Literature review and our experience during COVID-19 pandemic. *World Journal of Clinical Cases*, 11(29):6974–6983, October 2023. ISSN 2307-8960. doi: 10.12998/wjcc.v11.i29.6974. URL https://www.wjgnet.com/2307-8960/full/v11/i29/6974.htm.

[396] Han-Yi Robert Chiu, Chun-Kai Hwang, Shey-Ying Chen, Fuh-Yuan Shih, Hsieh-Cheng Han, Chwan-Chuen King, John Reuben Gilbert, Cheng-Chung Fang, and Yen-Jen Oyang. Machine learning for emerging infectious disease field responses. *Scientific Reports*, 12(1):328, January 2022. ISSN 2045-2322. doi: 10.1038/s41598-021-03687-w. URL https://www.nature.com/articles/s41598-021-03687-w.

[397] Firuz Kamalov, Khairan Rajab, Aswani Kumar Cherukuri, Ashraf Elnagar, and Murodbek Safaraliev. Deep learning for Covid-19 forecasting: State-of-the-art review. *Neurocomputing*, 511:142–154, October 2022. ISSN 09252312. doi: 10.1016/j.neucom.2022.09.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0925231222010918.

[398] Christopher M. Bishop and Hugh Bishop. *Deep learning: foundations and concepts*. Springer, Cham, Switzerland, 2024. ISBN 978-3-031-45467-7.

[399] Tom M. Mitchell. *Machine learning*. McGraw-Hill series in Computer Science. McGraw-Hill, New York, nachdr. edition, 2013. ISBN 978-0-07-042807-2 978-0-07-115467-3.

[400] Jose M. G. Vilar and Leonor Saiz. Dynamics-informed deconvolutional neural networks for super-resolution identification of regime changes in epidemiological time series. *Science Advances*, 9(28):eadf0673, July 2023. ISSN 2375-2548. doi: 10.1126/sciadv.adf0673. URL https://www.science.org/doi/10.1126/sciadv.adf0673.

[401] Azul Garza, Cristian Challu, and Max Mergenthaler-Canseco. TimeGPT-1, May 2024. URL http://arxiv.org/abs/2310.03589. arXiv:2310.03589 [cs].

[402] Karthik Gangavarapu, Alaa Abdel Latif, Julia L. Mullen, Manar Alkuzweny, Emory Hufbauer, Ginger Tsueng, Emily Haag, Mark Zeller, Christine M. Aceves, Karina Zaiets, Marco Cano, Jerry Zhou, Zhongchao Qian, Rachel Sattler, Nathaniel L. Matteson, Joshua I. Levy, Raphael Tc Lee, Lucas Freitas, Sebastian Maurer-Stroh, GISAID core and curation team, Marc A. Suchard, Chunlei Wu, Andrew I. Su, Kristian G. Andersen, and Laura D. Hughes. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Research Square*, pages rs.3.rs–1723829, June 2022. ISSN 2693-5015. doi: 10.21203/rs.3.rs-1723829/v1.

[403] Charles Nicholson, Lex Beattie, Matthew Beattie, Talayeh Razzaghi, and Sixia Chen. A machine learning and clustering-based approach for county-level COVID-19 analysis. *PloS One*, 17(4):e0267558, 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0267558.

[404] Rayner Alfred and Joe Henry Obit. The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review. *Heliyon*, 7(6):e07371, June 2021. ISSN 24058440. doi: 10.1016/j.heliyon.2021.e07371. URL https://linkinghub.elsevier.com/retrieve/pii/S2405844021014742.

[405] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986. ISSN 1476-4687. doi: 10.1038/323533a0. URL https://www.nature.com/articles/323533a0. Publisher: Nature Publishing Group.

[406] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, Cambridge, 2014. ISBN 978-1-107-05713-5. doi: 10.1017/CBO9781107298019. URL https://www.cambridge.org/core/books/understanding-machine-learning/3059695661405D25673058E43C8BE2A6.

[407] Ivan Noverlianto Tanawi, Valentino Vito, Devvi Sarwinda, Hengki Tasman, and Gatot Fatwanto Hertono. Support Vector Regression for Predicting the Number of Dengue Incidents in DKI Jakarta. *Procedia Computer Science*, 179: 747–753, 2021. ISSN 18770509. doi: 10.1016/j.procs.2021.01.063. URL https://linkinghub.elsevier.com/retrieve/pii/S187705092100079X.

[408] Xing Yu Leung, Rakibul M. Islam, Mohammadmehdi Adhami, Dragan Ilic, Lara McDonald, Shanika Palawaththa, Basia Diug, Saif U. Munshi, and Md Nazmul Karim. A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLOS Neglected Tropical Diseases*, 17(2):e0010631, February 2023. ISSN 1935-2735. doi: 10.1371/journal.pntd.0010631. URL https://dx.plos.org/10.1371/journal.pntd.0010631.

[409] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks, February 2018. URL http://arxiv.org/abs/1710.10903. arXiv:1710.10903 [cs, stat].

[410] Songgaojun Deng, Shusen Wang, Huzefa Rangwala, Lijing Wang, and Yue Ning. Cola-GNN: Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 245–254, Virtual Event Ireland, October 2020. ACM. ISBN 978-1-4503-6859-9. doi: 10.1145/3340531.3411975. URL https://dl.acm.org/doi/10.1145/3340531.3411975.

[411] Zewen Liu, Guancheng Wan, B. Aditya Prakash, Max S. Y. Lau, and Wei Jin. A Review of Graph Neural Networks in Epidemic Modeling, September 2024. URL http://arxiv.org/abs/2403.19852. arXiv:2403.19852 [cs].

[412] Soo Yong Lee, Fanchen Bu, Jaemin Yoo, and Kijung Shin. Towards Deep Attention in Graph Neural Networks: Problems and Remedies, June 2023. URL http://arxiv.org/abs/2306.02376. arXiv:2306.02376 [cs].

[413] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99:101805, November 2023. ISSN 15662535. doi: 10.1016/j.inffus.2023.101805. URL https://linkinghub.elsevier.com/retrieve/pii/S1566253523001148.

[414] Steven L. Brunton and J. Nathan Kutz. Promising directions of machine learning for partial differential equations. *Nature Computational Science*, 4(7):483–494, June 2024. ISSN 2662-8457. doi: 10.1038/s43588-024-00643-2. URL https://www.nature.com/articles/s43588-024-00643-2.

[415] Kate Crawford. Generative AI's environmental costs are soaring — and mostly secret. *Nature*, 626(8000):693–693, February 2024. doi: 10.1038/d41586-024-00478-x. URL https://www.nature.com/articles/d41586-024-00478-x. Bandiera_abtest: a Cg_type: World View Publisher: Nature Publishing Group Subject_term: Machine learning, Computer science, Technology, Policy.

[416] Pengfei Li, Jianyi Yang, Mohammad A. Islam, and Shaolei Ren. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models, March 2025. URL http://arxiv.org/abs/2304.03271. arXiv:2304.03271 [cs].

[417] Giulia Vilone and Luca Longo. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76:89–106, December 2021. ISSN 15662535. doi: 10.1016/j.inffus.2021.05.009. URL https://linkinghub.elsevier.com/retrieve/pii/S1566253521001093.

[418] Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions, November 2017. URL http://arxiv.org/abs/1705.07874. arXiv:1705.07874 [cs].

[419] Helen Coupland, Neil Scheidwasser, Alexandros Katsiferis, Megan Davies, Seth Flaxman, Naja Hulvej Rod, Swapnil Mishra, Samir Bhatt, and H. Juliette T. Unwin. Exploring the Potential and Limitations of Deep Learning and Explainable AI for Longitudinal Life Course Analysis, March 2025. URL https://www.researchsquare.com/article/rs-5582136/v1. ISSN: 2693-5015.

[420] Raha Moraffah, Mansooreh Karami, Ruocheng Guo, Adrienne Raglin, and Huan Liu. Causal Interpretability for Machine Learning – Problems, Methods and Evaluation, March 2020. URL http://arxiv.org/abs/2003.03934. arXiv:2003.03934 [cs].

[421] J.L. Ellis, M. Jacobs, J. Dijkstra, H. Van Laar, J.P. Cant, D. Tulpan, and N. Ferguson. Review: Synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data. *Animal*, 14:s223–s237, 2020. ISSN 17517311. doi: 10.1017/S1751731120000312. URL https://linkinghub.elsevier.com/retrieve/pii/S1751731120000312.

[422] Freddie Bickford Smith, Jannik Kossen, Eleanor Trollope, Mark van der Wilk, Adam Foster, and Tom Rainforth. Rethinking Aleatoric and Epistemic Uncertainty, 2024. URL https://arxiv.org/abs/2412.20892. Version Number: 1.

[423] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, 2015. URL https://arxiv.org/abs/1506.02142. Version Number: 6.

[424] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, 2016. URL https://arxiv.org/abs/1612.01474. Version Number: 3.

[425] Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification, 2021. URL https://arxiv.org/abs/2107.07511. Version Number: 6.

[426] Anastasios N. Angelopoulos, Emmanuel J. Candes, and Ryan J. Tibshirani. Conformal PID Control for Time Series Prediction, July 2023. URL http://arxiv.org/abs/2307.16895. arXiv:2307.16895 [cs, eess, stat].

[427] Liza Hadley, Caylyn Rich, Alex Tasker, Olivier Restif, and Sebastian Funk. How does policy modelling work in practice? A global analysis on the use of epidemiological modelling in health crises, August 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.08.12.24311899.

[428] Jasmina Panovska-Griffiths, Nicholas A. Watkins, Fergus Cumming, Luke Hounsome, Andre Charlett, Xu-Sheng Zhang, Thomas Finnie, Thomas Ward, Meera Chand, and Johanna Hutchinson. Responsive modelling of the mpox epidemic in England as a cross-government and academia collaboration: the process, challenges, and lessons learned. *The Lancet Microbe*, 4(9):e658–e660, September 2023. ISSN 2666-5247. doi: 10.1016/S2666-5247(23)00151-9. URL https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(23)00151-9/fulltext. Publisher: Elsevier.

[429] Mirjam E. Kretzschmar, Ben Ashby, Elizabeth Fearon, Christopher E. Overton, Jasmina Panovska-Griffiths, Lorenzo Pellis, Matthew Quaife, Ganna Rozhnova, Francesca Scarabel, Helena B. Stage, Ben Swallow, Robin N. Thompson, Michael J. Tildesley, and Daniel Villela. Challenges for modelling interventions for future pandemics. *Epidemics*, 38: 100546, March 2022. ISSN 1755-4365. doi: 10.1016/j.epidem.2022.100546. URL https://www.sciencedirect.com/science/article/pii/S1755436522000081.

[430] Gemma Hammerton and Marcus R. Munafò. Causal inference with observational data: the need for triangulation of evidence. *Psychological Medicine*, 51(4):563–578, March 2021. ISSN 0033-2917. doi: 10.1017/S0033291720005127. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8020490/.

[431] Tilmann Gneiting, Adrian E. Raftery, Anton H. Westveld, and Tom Goldman. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review*, 133(5):1098–1118, May 2005. ISSN 1520-0493, 0027-0644. doi: 10.1175/MWR2904.1. URL http://journals.ametsoc.org/doi/10.1175/MWR2904.1.

[432] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45, 2006. ISSN 1531-636X. doi: 10.1109/MCAS.2006.1688199. URL http://ieeexplore.ieee.org/document/1688199/.

[433] Evan L. Ray, Logan C. Brooks, Jacob Bien, Matthew Biggerstaff, Nikos I. Bosse, Johannes Bracher, Estee Y. Cramer, Sebastian Funk, Aaron Gerding, Michael A. Johansson, Aaron Rumack, Yijin Wang, Martha Zorn, Ryan J. Tibshirani, and Nicholas G. Reich. Comparing trained and untrained probabilistic ensemble forecasts of COVID-19 cases and deaths in the United States. *International Journal of Forecasting*, 39(3):1366–1383, July 2023. ISSN 01692070. doi: 10.1016/j.ijforecast.2022.06.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0169207022000966.

[434] Katharine Sherratt, Hugo Gruson, Rok Grah, Helen Johnson, Rene Niehus, Bastian Prasse, Frank Sandmann, Jannik Deuschel, Daniel Wolffram, Sam Abbott, Alexander Ullrich, Graham Gibson, Evan L Ray, Nicholas G Reich, Daniel Sheldon, Yijin Wang, Nutcha Wattanachit, Lijing Wang, Jan Trnka, Guillaume Obozinski, Tao Sun, Dorina Thanou, Loic Pottier, Ekaterina Krymova, Jan H Meinke, Maria Vittoria Barbarossa, Neele Leithauser, Jan Mohring, Johanna Schneider, Jaroslaw Wlazlo, Jan Fuhrmann, Berit Lange, Isti Rodiah, Prasith Baccam, Heidi Gurung, Steven Stage, Bradley Suchoski, Jozef Budzinski, Robert Walraven, Inmaculada Villanueva, Vit Tucek, Martin Smid, Milan Zajicek, Cesar Perez Alvarez, Borja Reina, Nikos I Bosse, Sophie R Meakin, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dave Osthus, Pierfrancesco Alaimo Di Loro, Antonello Maruotti, Veronika Eclerova, Andrea Kraus, David Kraus, Lenka Pribylova, Bertsimas Dimitris, Michael Lingzhi Li, Soni Saksham, Jonas Dehning, Sebastian Mohr, Viola Priesemann, Grzegorz Redlarski, Benjamin Bejar, Giovanni Ardenghi, Nicola Parolini, Giovanni Ziarelli, Wolfgang Bock, Stefan Heyder, Thomas Hotz, David E Singh, Miguel Guzman-Merino, Jose L Aznarte, David Morina, Sergio Alonso, Enric Alvarez, Daniel Lopez, Clara Prats, Jan Pablo Burgard, Arne Rodloff, Tom Zimmermann, Alexander Kuhlmann, Janez Zibert, Fulvia Pennoni, Fabio Divino, Marti Catala, Gianfranco Lovison, Paolo Giudici, Barbara Tarantino, Francesco Bartolucci, Giovanna Jona Lasinio, Marco Mingione, Alessio Farcomeni, Ajitesh Srivastava, Pablo Montero-Manso, Aniruddha Adiga, Benjamin Hurt, Bryan Lewis, Madhav Marathe, Przemyslaw Porebski, Srinivasan Venkatramanan, Rafal P Bartczuk, Filip Dreger, Anna Gambin, Krzysztof Gogolewski, Magdalena Gruziel-Slomka, Bartosz Krupa, Antoni Moszyński, Karol Niedzielewski, Jedrzej Nowosielski, Maciej Radwan, Franciszek Rakowski, Marcin Semeniuk, Ewa Szczurek, Jakub Zielinski, Jan Kisielewski, Barbara Pabjan, Kirsten Holger, Yuri Kheifetz, Markus Scholz, Biecek Przemyslaw, Marcin Bodych, Maciej Filinski, Radoslaw Idzikowski, Tyll Krueger, Tomasz Ozanski, Johannes Bracher, and Sebastian Funk. Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. *eLife*, 12:e81916, April 2023. ISSN 2050-084X. doi: 10.7554/eLife.81916. URL https://elifesciences.org/articles/81916.

[435] C. Jessica E. Metcalf, Dylan H. Morris, and Sang Woo Park. Mathematical models to guide pandemic response. *Science*, 369(6502):368–369, July 2020. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abd1668. URL https://www.science.org/doi/10.1126/science.abd1668.

[436] Shimoni Shah, Sylvia Xiao Wei Gwee, Jamie Qiao Xin Ng, Nicholas Lau, Jiayun Koh, and Junxiong Pang. Wastewater surveillance to infer COVID-19 transmission: A systematic review. *Science of The Total Environment*, 804:150060, January 2022. ISSN 00489697. doi: 10.1016/j.scitotenv.2021.150060. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969721051354.

[437] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Muñoz-Marí, Egbert H. Van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Schölkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang, and Jakob Zscheischler. Inferring causation from time series in Earth system sciences. *Nature Communications*, 10(1): 2553, June 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-10105-3. URL https://www.nature.com/articles/s41467-019-10105-3.

[438] Cara E. Brook, Carly Rozins, Jennifer A. Bohl, Vida Ahyong, Sophana Chea, Liz Fahsbender, Rekol Huy, Sreyngim Lay, Rithea Leang, Yimei Li, Chanthap Lon, Somnang Man, Mengheng Oum, Graham R. Northrup, Fabiano Oliveira, Andrea R. Pacheco, Daniel M. Parker, Katherine Young, Michael Boots, Cristina M. Tato, Joseph L. DeRisi, Christina Yek, and Jessica E. Manning. Climate, demography, immunology, and virology combine to drive two decades of dengue virus dynamics in Cambodia, June 2022. URL http://medrxiv.org/lookup/doi/10.1101/2022.06.08.22276171.

[439] David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin

Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, August 2017. ISSN 0906-7590, 1600-0587. doi: 10.1111/ecog.02881. URL https://onlinelibrary.wiley.com/doi/10.1111/ecog.02881.

[440] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. Adaptive computation and machine learning. The MIT press, Cambridge, Mass, 2017. ISBN 978-0-262-03731-0.

[441] Karl J. Friston, Thomas Parr, Peter Zeidman, Adeel Razi, Guillaume Flandin, Jean Daunizeau, Ollie J. Hulme, Alexander J. Billig, Vladimir Litvak, Rosalyn J. Moran, Cathy J. Price, and Christian Lambert. Dynamic causal modelling of COVID-19. *Wellcome Open Research*, 5:89, August 2020. ISSN 2398-502X. doi: 10.12688/wellcomeopenres.15881.2. URL https://wellcomeopenresearch.org/articles/5-89/v2.

[442] Dinh Viet Cuong, Branislava Lalić, Mina Petrić, Binh Nguyen, and Mark Roantree. Adapting Physics-Informed Neural Networks To Optimize ODEs in Mosquito Population Dynamics, June 2024. URL http://arxiv.org/abs/2406.05108. arXiv:2406.05108 [cs, q-bio].

[443] Katriona Shea, Rebecca K. Borchering, William J. M. Probert, Emily Howerton, Tiffany L. Bogich, Shou-Li Li, Willem G. Van Panhuis, Cecile Viboud, Ricardo Aguás, Artur A. Belov, Sanjana H. Bhargava, Sean M. Cavany, Joshua C. Chang, Cynthia Chen, Jinghui Chen, Shi Chen, YangQuan Chen, Lauren M. Childs, Carson C. Chow, Isabel Crooker, Sara Y. Del Valle, Guido España, Geoffrey Fairchild, Richard C. Gerkin, Timothy C. Germann, Quanquan Gu, Xiangyang Guan, Lihong Guo, Gregory R. Hart, Thomas J. Hladish, Nathaniel Hupert, Daniel Janies, Cliff C. Kerr, Daniel J. Klein, Eili Y. Klein, Gary Lin, Carrie Manore, Lauren Ancel Meyers, John E. Mittler, Kunpeng Mu, Rafael C. Núñez, Rachel J. Oidtman, Remy Pasco, Ana Pastore Y Piontti, Rajib Paul, Carl A. B. Pearson, Dianela R. Perdomo, T. Alex Perkins, Kelly Pierce, Alexander N. Pillai, Rosalyn Cherie Rael, Katherine Rosenfeld, Chrysm Watson Ross, Julie A. Spencer, Arlin B. Stoltzfus, Kok Ben Toh, Shashaank Vattikuti, Alessandro Vespignani, Lingxiao Wang, Lisa J. White, Pan Xu, Yupeng Yang, Osman N. Yogurtcu, Weitong Zhang, Yanting Zhao, Difan Zou, Matthew J. Ferrari, David Pannell, Michael J. Tildesley, Jack Seifarth, Elyse Johnson, Matthew Biggerstaff, Michael A. Johansson, Rachel B. Slayton, John D. Levander, Jeff Stazer, Jessica Kerr, and Michael C. Runge. Multiple models for outbreak decision support in the face of uncertainty. *Proceedings of the National Academy of Sciences*, 120(18):e2207537120, May 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2207537120. URL https://pnas.org/doi/10.1073/pnas.2207537120.

[444] Mircea T. Sofonea, Simon Cauchemez, and Pierre-Yves Boëlle. Epidemic models: why and how to use them. *Anaesthesia, Critical Care & Pain Medicine*, 41(2):101048, April 2022. ISSN 2352-5568. doi: 10.1016/j.accpm.2022.101048.

[445] Anastasia Chatzilena, Edwin Van Leeuwen, Oliver Ratmann, Marc Baguelin, and Nikolaos Demiris. Contemporary statistical inference for infectious disease models using Stan. *Epidemics*, 29:100367, December 2019. ISSN 17554365. doi: 10.1016/j.epidem.2019.100367. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436519300325.

[446] Theodore Kypraios, Peter Neal, and Dennis Prangle. A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. *Mathematical Biosciences*, 287:42–53, May 2017. ISSN 0025-5564. doi: 10.1016/j.mbs.2016.07.001. URL https://www.sciencedirect.com/science/article/pii/S0025556416300839.

[447] J. L. Doob. Markoff Chains–Denumerable Case. *Transactions of the American Mathematical Society*, 58(3):455, November 1945. ISSN 00029947. doi: 10.2307/1990339. URL https://www.jstor.org/stable/1990339?origin=crossref.

[448] David Welch, Shweta Bansal, and David R. Hunter. Statistical inference to advance network models in epidemiology. *Epidemics*, 3(1):38–45, March 2011. ISSN 1878-0067. doi: 10.1016/j.epidem.2011.01.002.

[449] Niels G. Becker. Parametric inference for epidemic models. *Mathematical Biosciences*, 117(1-2):239–251, September 1993. ISSN 00255564. doi: 10.1016/0025-5564(93)90026-7. URL https://linkinghub.elsevier.com/retrieve/pii/0025556493900267.

[450] Arnaud Doucet, Nando Freitas, and Neil Gordon. An Introduction to Sequential Monte Carlo Methods. In Arnaud Doucet, Nando Freitas, and Neil Gordon, editors, *Sequential Monte Carlo Methods in Practice*, pages 3–14. Springer New York, New York, NY, 2001. ISBN 978-1-4419-2887-0 978-1-4757-3437-9. doi: 10.1007/978-1-4757-3437-9_1. URL http://link.springer.com/10.1007/978-1-4757-3437-9_1.

[451] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. 2017. doi: 10.48550/ARXIV.1701.02434. URL https://arxiv.org/abs/1701.02434. Publisher: arXiv Version Number: 2.

[452] George Casella and Edward I. George. Explaining the Gibbs Sampler. *The American Statistician*, 46(3):167–174, August 1992. ISSN 0003-1305, 1537-2731. doi: 10.1080/00031305.1992.10475878. URL http://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475878.

[453] Radford M. Neal. *MCMC using Hamiltonian dynamics*. May 2011. doi: 10.1201/b10905. URL http://arxiv.org/abs/1206.1901. arXiv:1206.1901 [stat].

[454] Nicholas Steyn, Kris V. Parag, Robin N. Thompson, and Christl A. Donnelly. A primer on inference and prediction with epidemic renewal models and sequential Monte Carlo, March 2025. URL http://arxiv.org/abs/2503.18875. arXiv:2503.18875 [stat].

[455] Aaron A. King, Edward L. Ionides, Carles Martinez Bretó, Stephen P. Ellner, Matthew J. Ferrari, Sebastian Funk, Steven G. Johnson, Bruce E. Kendall, Michael Lavine, Dao Nguyen, Eamon B. O'Dea, Daniel C. Reuman, Helen Wearing, and Simon N. Wood. *pomp: Statistical Inference for Partially Observed Markov Processes*. 2024. URL https://kingaa.github.io/pomp/.

[456] Kidus Asfaw, Joonha Park, Aaron A. King, and Edward L. Ionides. A tutorial on spatiotemporal partially observed Markov process models via the R package spatPomp, April 2024. URL http://arxiv.org/abs/2101.01157. arXiv:2101.01157 [stat].

[457] Amanda Minter and Renata Retkute. Approximate Bayesian Computation for infectious disease modelling. *Epidemics*, 29:100368, December 2019. ISSN 1878-0067. doi: 10.1016/j.epidem.2019.100368.

[458] Stan. RStan: the R interface to Stan, 2023. URL https://mc-stan.org/.

[459] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392, April 2009. ISSN 1369-7412, 1467-9868. doi: 10.1111/j.1467-9868.2008.00700.x. URL https://academic.oup.com/jrsssb/article/71/2/319/7092907.

[460] Haakon Bakka, Håvard Rue, Geir-Arne Fuglstad, Andrea Riebler, David Bolin, Janine Illian, Elias Krainski, Daniel Simpson, and Finn Lindgren. Spatial modeling with R-INLA: A review. *WIREs Computational Statistics*, 10(6):e1443, November 2018. ISSN 1939-5108, 1939-0068. doi: 10.1002/wics.1443. URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1443.

[461] Janet Van Niekerk, Elias Krainski, Denis Rustand, and Håvard Rue. A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, 181:107692, May 2023. ISSN 01679473. doi: 10.1016/j.csda.2023.107692. URL https://linkinghub.elsevier.com/retrieve/pii/S0167947323000038.

[462] Sebastian Ruder. An overview of gradient descent optimization algorithms, 2016. URL https://arxiv.org/abs/1609.04747. Version Number: 2.

[463] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, 2014. URL https://arxiv.org/abs/1412.6980. Version Number: 9.

[464] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716–723, December 1974. ISSN 0018-9286. doi: 10.1109/TAC.1974.1100705. URL http://ieeexplore.ieee.org/document/1100705/.

[465] Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), March 1978. ISSN 0090-5364. doi: 10.1214/aos/1176344136. URL https://projecteuclid.org/journals/annals-of-statistics/volume-6/issue-2/Estimating-the-Dimension-of-a-Model/10.1214/aos/1176344136.full.

[466] Aki Vehtari, Jonah Gabry, Mans Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models, 2024. URL https://mc-stan.org/loo/.

[467] Gerardo Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, August 2017. ISSN 2468-0427. doi: 10.1016/j.idm.2017.08.001. URL https://www.sciencedirect.com/science/article/pii/S2468042717300234.

[468] Farzaneh Sadat Tabataba, Prithwish Chakraborty, Naren Ramakrishnan, Srinivasan Venkatramanan, Jiangzhuo Chen, Bryan Lewis, and Madhav Marathe. A framework for evaluating epidemic forecasts. *BMC Infectious Diseases*, 17(1):345, December 2017. ISSN 1471-2334. doi: 10.1186/s12879-017-2365-1. URL http://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-017-2365-1.

[469] Daniel Wolffram, Sam Abbott, Matthias an der Heiden, Sebastian Funk, Felix Günther, Davide Hailer, Stefan Heyder, Thomas Hotz, Jan van de Kassteele, Helmut Küchenhoff, Sören Müller-Hansen, Diellë Syliqi, Alexander Ullrich, Maximilian Weigert, Melanie Schienle, and Johannes Bracher. Collaborative nowcasting of COVID-19 hospitalization incidences in Germany. *PLOS Computational Biology*, 19(8):e1011394, August 2023. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011394. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011394. Publisher: Public Library of Science.

[470] Tilmann Gneiting and Adrian E Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, March 2007. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214506000001437. URL http://www.tandfonline.com/doi/abs/10.1198/016214506000001437.

[471] Daniela De Angelis, Anne M. Presanis, Paul J. Birrell, Gianpaolo Scalia Tomba, and Thomas House. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*, 10:83–87, March 2015. ISSN 17554365. doi: 10.1016/j.epidem.2014.09.004. URL https://linkinghub.elsevier.com/retrieve/pii/S175543651400053X.

[472] Pragati V. Prasad, Molly K. Steele, Carrie Reed, Lauren Ancel Meyers, Zhanwei Du, Remy Pasco, Jorge A. Alfaro-Murillo, Bryan Lewis, Srinivasan Venkatramanan, James Schlitt, Jiangzhuo Chen, Mark Orr, Mandy L. Wilson, Stephen Eubank, Lijing Wang, Matteo Chinazzi, Ana Pastore y Piontti, Jessica T. Davis, M. Elizabeth Halloran, Ira Longini, Alessandro Vespignani, Sen Pei, Marta Galanti, Sasikiran Kandula, Jeffrey Shaman, David J. Haw, Nimalan Arinaminpathy, and Matthew Biggerstaff. Multimodeling approach to evaluating the efficacy of layering pharmaceutical and nonpharmaceutical interventions for influenza pandemics. *Proceedings of the National Academy of Sciences*, 120 (28):e2300590120, July 2023. doi: 10.1073/pnas.2300590120. URL https://www.pnas.org/doi/10.1073/pnas.2300590120. Publisher: Proceedings of the National Academy of Sciences.

[473] Evan L. Ray, Yijin Wang, Russell D. Wolfinger, and Nicholas G. Reich. Flusion: Integrating multiple data sources for accurate influenza predictions. *Epidemics*, page 100810, December 2024. ISSN 17554365. doi: 10.1016/j.epidem.2024.100810. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436524000719.

[474] Jennifer Weidhaas, Zachary T. Aanderud, D. Keith Roper, James VanDerslice, Erica Brown Gaddis, Jeff Ostermiller, Ken Hoffman, Rubayat Jamal, Phillip Heck, Yue Zhang, Kevin Torgersen, Jacob Vander Laan, and Nathan LaCross. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Science of The Total Environment*, 775:145790, June 2021. ISSN 00489697. doi: 10.1016/j.scitotenv.2021.145790. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969721008573.

[475] Doug Manuel, Carlo Alberto Amadei, Jonathon R Campbell, Jean-Martin Brault, and Jeremy Veillard. Strengthening public health surveillance through wastewater testing. 2022. Publisher: World Bank, Washington, DC.

[476] Gertjan Medema, Leo Heijnen, Goffe Elsinga, Ronald Italiaander, and Anke Brouwer. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters*, 7(7):511–516, July 2020. ISSN 2328-8930, 2328-8930. doi: 10.1021/acs.estlett.0c00357. URL https://pubs.acs.org/doi/10.1021/acs.estlett.0c00357.

[477] Jordan Peccia, Alessandro Zulli, Doug E. Brackney, Nathan D. Grubaugh, Edward H. Kaplan, Arnau Casanovas-Massana, Albert I. Ko, Amyn A. Malik, Dennis Wang, Mike Wang, Joshua L. Warren, Daniel M. Weinberger, Wyatt Arnold, and Saad B. Omer. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology*, 38(10):1164–1167, October 2020. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-020-0684-z. URL https://www.nature.com/articles/s41587-020-0684-z.

[478] Amy Xiao, Fuqing Wu, Mary Bushman, Jianbo Zhang, Maxim Imakaev, Peter R Chai, Claire Duvallet, Noriko Endo, Timothy B Erickson, Federica Armas, Brian Arnold, Hongjie Chen, Franciscus Chandra, Newsha Ghaeli, Xiaoqiong Gu, William P Hanage, Wei Lin Lee, Mariana Matus, Kyle A McElroy, Katya Moniz, Steven F Rhode, Janelle Thompson, and Eric J Alm. Metrics to relate COVID-19 wastewater data to clinical testing dynamics. *Water Research*, 212:118070, April 2022. ISSN 00431354. doi: 10.1016/j.watres.2022.118070. URL https://linkinghub.elsevier.com/retrieve/pii/S0043135422000331.

[479] Mario Morvan, Anna Lo Jacomo, Celia Souque, Matthew J. Wade, Till Hoffmann, Koen Pouwels, Chris Lilley, Andrew C. Singer, Jonathan Porter, Nicholas P. Evens, David I. Walker, Joshua T. Bunce, Andrew Engeli, Jasmine

Grimsley, Kathleen M. O'Reilly, and Leon Danon. An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. *Nature Communications*, 13(1):4313, July 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-31753-y. URL https://www.nature.com/articles/s41467-022-31753-y.

[480] Guangquan Li, Hubert Denise, Peter Diggle, Jasmine Grimsley, Chris Holmes, Daniel James, Radka Jersakova, Callum Mole, George Nicholson, Camila Rangel Smith, Sylvia Richardson, William Rowe, Barry Rowlingson, Fatemeh Torabi, Matthew J. Wade, and Marta Blangiardo. A spatio-temporal framework for modelling wastewater concentration during the COVID-19 pandemic. *Environment International*, 172:107765, February 2023. ISSN 01604120. doi: 10.1016/j.envint.2023.107765. URL https://linkinghub.elsevier.com/retrieve/pii/S0160412023000387.

[481] Guangquan Li, Peter Diggle, and Marta Blangiardo. Integrating wastewater and randomised prevalence survey data for national COVID surveillance. *Scientific Reports*, 14(1):5124, March 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-55752-9. URL https://www.nature.com/articles/s41598-024-55752-9.

[482] C. J. Clopper and E. S. Pearson. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26(4):404–413, 1934. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/26.4.404. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/26.4.404.

[483] Kei Miyakawa, Masakazu Machida, Tomoyuki Kawasaki, Mayuko Nishi, Hidenori Akutsu, and Akihide Ryo. Reduced Replication Efficacy of Severe Acute Respiratory Syndrome Coronavirus 2 Omicron Variant in "Mini-gut" Organoids. *Gastroenterology*, 163(2):514–516, August 2022. ISSN 00165085. doi: 10.1053/j.gastro.2022.04.043. URL https://linkinghub.elsevier.com/retrieve/pii/S0016508522004565.

[484] Célia Sentis, Geneviève Billaud, Antonin Bal, Emilie Frobert, Maude Bouscambert, Gregory Destras, Laurence Josset, Bruno Lina, Florence Morfin, Alexandre Gaymard, and the COVID-Diagnosis HCL Study Group. SARS-CoV-2 Omicron Variant, Lineage BA.1, Is Associated with Lower Viral Load in Nasopharyngeal Samples Compared to Delta Variant. *Viruses*, 14(5):919, April 2022. ISSN 1999-4915. doi: 10.3390/v14050919. URL https://www.mdpi.com/1999-4915/14/5/919.

[485] Young R Jang, Jeong-Min Kim, Jee E Rhee, Dongju Kim, Nam-Joo Lee, Hyeokjin Lee, Jong-Hun Kim, Eun-Jin Kim, and Jin Y Kim. Clinical Features and Duration of Viral Shedding in Individuals With SARS-CoV-2 Omicron Variant Infection. *Open Forum Infectious Diseases*, 9(7):ofac237, July 2022. ISSN 2328-8957. doi: 10.1093/ofid/ofac237. URL https://academic.oup.com/ofid/article/doi/10.1093/ofid/ofac237/6582525.

[486] Paul Elliott, Oliver Eales, Barbara Bodinier, David Tang, Haowei Wang, Jakob Jonnerby, David Haw, Joshua Elliott, Matthew Whitaker, Caroline E. Walters, Christina Atchison, Peter J. Diggle, Andrew J. Page, Alexander J. Trotter, Deborah Ashby, Wendy Barclay, Graham Taylor, Helen Ward, Ara Darzi, Graham S. Cooke, Marc Chadeau-Hyam, and Christl A. Donnelly. Dynamics of a national Omicron SARS-CoV-2 epidemic during January 2022 in England. *Nature Communications*, 13(1):4500, August 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32121-6. URL https://www.nature.com/articles/s41467-022-32121-6.

[487] Aaron Bivins and Kyle Bibby. Wastewater Surveillance during Mass COVID-19 Vaccination on a College Campus. *Environmental Science & Technology Letters*, 8(9):792–798, September 2021. ISSN 2328-8930, 2328-8930. doi: 10.1021/acs.estlett.1c00519. URL https://pubs.acs.org/doi/10.1021/acs.estlett.1c00519.

[488] Kenrie P. Y. Hui, John C. W. Ho, Man-chun Cheung, Ka-chun Ng, Rachel H. H. Ching, Ka-ling Lai, Tonia Tong Kam, Haogao Gu, Ko-Yung Sit, Michael K. Y. Hsin, Timmy W. K. Au, Leo L. M. Poon, Malik Peiris, John M. Nicholls, and Michael C. W. Chan. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo.

Nature, 603(7902):715–720, March 2022. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-022-04479-6. URL https://www.nature.com/articles/s41586-022-04479-6.

[489] Matthew J. Wade, Anna Lo Jacomo, Elena Armenise, Mathew R. Brown, Joshua T. Bunce, Graeme J. Cameron, Zhou Fang, Kata Farkas, Deidre F. Gilpin, David W. Graham, Jasmine M.S. Grimsley, Alwyn Hart, Till Hoffmann, Katherine J. Jackson, David L. Jones, Chris J. Lilley, John W. McGrath, Jennifer M. McKinley, Cormac McSparron, Behnam F. Nejad, Mario Morvan, Marcos Quintela-Baluja, Adrian M.I. Roberts, Andrew C. Singer, Célia Souque, Vanessa L. Speight, Chris Sweetapple, David Walker, Glenn Watts, Andrew Weightman, and Barbara Kasprzyk-Hordern. Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. *Journal of Hazardous Materials*, 424:127456, February 2022. ISSN 03043894. doi: 10.1016/j.jhazmat.2021.127456. URL https://linkinghub.elsevier.com/retrieve/pii/S0304389421024249.

[490] Olga E. Hart and Rolf U. Halden. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Science of The Total Environment*, 730:138875, August 2020. ISSN 00489697. doi: 10.1016/j.scitotenv.2020.138875. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969720323925.

[491] Manish Kumar, Arbind Kumar Patel, Anil V. Shah, Janvi Raval, Neha Rajpara, Madhvi Joshi, and Chaitanya G. Joshi. First proof of the capability of wastewater surveillance for COVID-19 in India through detection of genetic material of SARS-CoV-2. *Science of The Total Environment*, 746:141326, December 2020. ISSN 00489697. doi: 10.1016/j.scitotenv.2020.141326. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969720348555.

[492] Trevor Sharot. Weighting survey results. *J Mark Res Soc*, 28(3):269–284, 1986.

[493] Till Hoffmann, Sarah Bunney, Barbara Kasprzyk-Hordern, and Andrew C. Singer. Wastewater catchment areas in Great Britain. November 2022. doi: 10.1002/essoar.10510612.1.

[494] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. 2016. doi: 10.48550/ARXIV.1603.02754. URL https://arxiv.org/abs/1603.02754. Publisher: arXiv Version Number: 3.

[495] R Core Team. R: A Language and Environment for Statistical Computing, 2022. URL https://www.R-project.org/.

[496] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29 (5), October 2001. ISSN 0090-5364. doi: 10.1214/aos/1013203451. URL https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full.

[497] Jing Liu-Helmersson, Hans Stenlund, Annelies Wilder-Smith, and Joacim Rocklöv. Vectorial Capacity of Aedes aegypti: Effects of Temperature and Implications for Global Dengue Epidemic Potential. *PLoS ONE*, 9(3):e89783, March 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0089783. URL https://dx.plos.org/10.1371/journal.pone.0089783.

[498] Annabel Seah, Joel Aik, Lee-Ching Ng, and Clarence C. Tam. The effects of maximum ambient temperature and heatwaves on dengue infections in the tropical city-state of Singapore – A time series analysis. *Science of The Total Environment*, 775:145117, June 2021. ISSN 00489697. doi: 10.1016/j.scitotenv.2021.145117. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969721001832.

[499] Alexandre S. Gagnon, Andrew B. G. Bush, and Karen E. Smoyer-Tomic. Dengue epidemics and the El Niño Southern Oscillation. *Climate Research*, 19(1):35–43, 2001. ISSN 0936577X, 16161572. URL http://www.jstor.org/stable/24866766. Publisher: Inter-Research Science Center.

[500] Chuanxi Li, Zhao Liu, Wen Li, Yuxi Lin, Liangyu Hou, Shuyue Niu, Yue Xing, Jianbin Huang, Yidan Chen, Shangchen Zhang, Xuejie Gao, Ying Xu, Can Wang, Qi Zhao, Qiyong Liu, Wei Ma, Wenjia Cai, Peng Gong, and Yong Luo. Projecting future risk of dengue related to hydrometeorological conditions in mainland China under climate change scenarios: a modelling study. *The Lancet Planetary Health*, 7(5):e397–e406, May 2023. ISSN 25425196. doi: 10.1016/S2542-5196(23)00051-7. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519623000517.

[501] Azael Che-Mendoza, Abdiel Martin-Park, Juan Manuel Chávez-Trava, Yamili Contreras-Perera, Hugo Delfín-González, Gabriela González-Olvera, Jorge Leirana-Alcocer, Guillermo Guillermo-May, Daniel Chan-Espinoza, Norma Pavia-Ruz, Rosa Eugenia Méndez-Vales, Alberto Alcocer-Gamboa, Fabian Correa-Morales, Jorge Palacio-Vargas, Dongjing Zhang, Gonzalo Vazquez-Prokopec, Zhiyong Xi, and Pablo Manrique-Saide. Abundance and Seasonality of *Aedes aegypti* (Diptera: Culicidae) in Two Suburban Localities of South Mexico, With Implications for *Wolbachia* (Rickettsiales: Rickettsiaceae)-Carrying Male Releases for Population Suppression. *Journal of Medical Entomology*, 58(4):1817–1825, July 2021. ISSN 0022-2585, 1938-2928. doi: 10.1093/jme/tjab052. URL https://academic.oup.com/jme/article/58/4/1817/6208904.

[502] Bingyi Yang, Brooke A. Borgert, Barry W. Alto, Carl K. Boohene, Joe Brew, Kelly Deutsch, James T. DeValerio, Rhoel R. Dinglasan, Daniel Dixon, Joseph M. Faella, Sandra L. Fisher-Grainger, Gregory E. Glass, Reginald Hayes, David F. Hoel, Austin Horton, Agne Janusauskaite, Bill Kellner, Moritz U. G. Kraemer, Keira J. Lucas, Johana Medina, Rachel Morreale, William Petrie, Robert C. Reiner, Michael T. Riles, Henrik Salje, David L. Smith, John P. Smith, Amy Solis, Jason Stuck, Chalmers Vasquez, Katie F. Williams, Rui-De Xue, and Derek A. T. Cummings. Modelling distributions of Aedes aegypti and Aedes albopictus using climate, host density and interspecies competition. *PLOS Neglected Tropical Diseases*, 15(3):e0009063, March 2021. ISSN 1935-2735. doi: 10.1371/journal.pntd.0009063. URL https://dx.plos.org/10.1371/journal.pntd.0009063.

[503] Nur Athen Mohd Hardy Abdullah, Nazri Che Dom, Siti Aekball Salleh, Hasber Salim, and Nopadol Precha. The association between dengue case and climate: A systematic review and meta-analysis. *One Health*, 15:100452, December 2022. ISSN 23527714. doi: 10.1016/j.onehlt.2022.100452. URL https://linkinghub.elsevier.com/retrieve/pii/S2352771422000842.

[504] César V. Munayco, Betsabet Yadira Valderrama Rosales, Susan Yanett Mateo Lizarbe, Carmen Rosa Yon Fabian, Ricardo Peña Sánchez, César Henry Vásquez Sánchez, Maria Paquita García, Carlos Padilla-Rojas, Victor Suárez, Liliana Sánchez-González, Forrest K. Jones, Luciana Kohatsu, Laura E. Adams, Juliette Morgan, and Gabriela Paz-Bailey. *Notes from the Field:* Dengue Outbreak — Peru, 2023. *MMWR. Morbidity and Mortality Weekly Report*, 73(4):86–88, February 2024. ISSN 0149-2195, 1545-861X. doi: 10.15585/mmwr.mm7304a4. URL http://www.cdc.gov/mmwr/volumes/73/wr/mm7304a4.htm?s_cid=mm7304a4_w.

[505] CDC Peru. Reporte de Tabla de casos notificados por causas, May 2023. URL https://www.dge.gob.pe/salasituacional/.

[506] INEI. Estimated Population as of June 30, by calendar year and sex, by department, 2023. URL https://www.inei.gob.pe/media/MenuRecursivo/indices_tematicos/proy_04_3.xls.

[507] Tim Appelhans, Florian Detsch, Christoph Reudenbach, and Stefan Woellauer. mapview: Interactive Viewing of Spatial Data in R, December 2015. URL https://CRAN.R-project.org/package=mapview. Institution: Comprehensive R Archive Network Pages: 2.11.2.

[508] Daniel Runfola, Austin Anderson, Heather Baier, Matt Crittenden, Elizabeth Dowker, Sydney Fuhrig, Seth Goodman, Grace Grimsley, Rachel Layko, Graham Melville, Maddy Mulder, Rachel Oberman, Joshua Panganiban, Andrew Peck, Leigh Seitz, Sylvia Shea, Hannah Slevin, Rebecca Youngerman, and Lauren Hobbs. geoBoundaries: A global database of political administrative boundaries. *PLOS ONE*, 15(4):e0231866, April 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0231866. URL https://dx.plos.org/10.1371/journal.pone.0231866.

[509] Ian Harris, Timothy J. Osborn, Phil Jones, and David Lister. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1):109, April 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-0453-3. URL https://www.nature.com/articles/s41597-020-0453-3.

[510] B. Gruvberger, M. Bruze, and M. Tammela. Preservatives in moisturizers on the Swedish market. *Acta Dermato-Venereologica*, 78(1):52–56, January 1998. ISSN 0001-5555. doi: 10.1080/00015559850135850.

[511] Thomas B McKee, Nolan J Doesken, John Kleist, and others. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. California, 1993. Issue: 22.

[512] European Drought Observatory. Standardized Precipitation Index (SPI), 2023. URL https://edo.jrc.ec.europa.eu/gdo/php/index.php?id=2112.

[513] K. Takahashi, A. Montecinos, K. Goubanova, and B. Dewitte. ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophysical Research Letters*, 38(10):n/a–n/a, May 2011. ISSN 00948276. doi: 10.1029/2011GL047364. URL http://doi.wiley.com/10.1029/2011GL047364.

[514] Ken Takahashi, Kobi Mosquera, and Jorge Reupo. The El Niño Coastal Index (ICEN): history and update. Technical bulletin, Geophysical Institute of Peru, February 2014.

[515] NOAA. Cold & Warm Episodes by Season, July 2023. URL https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.

[516] Jan Verbesselt, Rob Hyndman, Glenn Newnham, and Darius Culvenor. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114(1):106–115, January 2010. ISSN 00344257. doi: 10.1016/j.rse.2009.08.014. URL https://linkinghub.elsevier.com/retrieve/pii/S003442570900265X.

[517] Jan Verbesselt, Rob Hyndman, Achim Zeileis, and Darius Culvenor. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12):2970–2980, December 2010. ISSN 00344257. doi: 10.1016/j.rse.2010.08.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0034425710002336.

[518] J. Welles Wilder. *New concepts in technical trading systems*. Trend Research, Greensboro, N.C, 1978. ISBN 978-0-89459-027-6.

[519] Julian Besag, Jeremy York, and Annie Mollie. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, March 1991. ISSN 0020-3157, 1572-9052. doi: 10.1007/BF00116466. URL http://link.springer.com/10.1007/BF00116466.

[520] Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1), February 2017. ISSN 0883-4237. doi: 10.1214/16-STS576. URL https://projecteuclid.org/journals/statistical-science/volume-32/issue-1/Penalising-Model-Component-Complexity--A-Principled-Practical-Approach-to/10.1214/16-STS576.full.

[521] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(4):583–639, October 2002. ISSN 1369-7412, 1467-9868. doi: 10.1111/1467-9868.00353. URL https://academic.oup.com/jrsssb/article/64/4/583/7098621.

[522] D. Collett. *Modelling binary data*. Chapman & Hall/CRC texts in statistical science series. Chapman & Hall/CRC, Boca Raton, 2nd ed edition, 2003. ISBN 978-1-58488-324-1.

[523] Khin Thet Wai, Natarajan Arunachalam, Susilowati Tana, Fe Espino, Pattamaporn Kittayapong, W Abeyewickreme, Dilini Hapangama, Brij Kishore Tyagi, Pe Than Htun, Surachart Koyadun, Axel Kroeger, Johannes Sommerfeld, and Max Petzold. Estimating dengue vector abundance in the wet and dry season: implications for targeted vector control in urban and peri-urban Asia. *Pathogens and Global Health*, 106(8):436–445, December 2012. ISSN 2047-7724, 2047-7732. doi: 10.1179/2047773212Y.0000000063. URL http://www.tandfonline.com/doi/full/10.1179/2047773212Y.0000000063.

[524] Lars Eisen, Julián E. García-Rejón, Salvador Gómez-Carro, María Del Rosario Nájera Vázquez, Thomas J. Keefe, Barry J. Beaty, and María Alba Loroño-Pino. Temporal Correlations Between Mosquito-Based Dengue Virus Surveillance Measures or Indoor Mosquito Abundance and Dengue Case Numbers in Mérida City, México. *Journal of Medical Entomology*, 51(4):885–890, July 2014. ISSN 0022-2585, 1938-2928. doi: 10.1603/ME14008. URL https://academic.oup.com/jme/article-lookup/doi/10.1603/ME14008.

[525] Shaowei Sang, Shaohua Gu, Peng Bi, Weizhong Yang, Zhicong Yang, Lei Xu, Jun Yang, Xiaobo Liu, Tong Jiang, Haixia Wu, Cordia Chu, and Qiyong Liu. Predicting Unprecedented Dengue Outbreak Using Imported Cases and Climatic Factors in Guangzhou, 2014. *PLOS Neglected Tropical Diseases*, 9(5):e0003808, May 2015. ISSN 1935-2735. doi: 10.1371/journal.pntd.0003808. URL https://dx.plos.org/10.1371/journal.pntd.0003808.

[526] L. Brown, J. Medlock, and V. Murray. Impact of drought on vector-borne diseases–how does one manage the risk? *Public Health*, 128(1):29–37, January 2014. ISSN 1476-5616. doi: 10.1016/j.puhe.2013.09.006.

[527] Brendan J. Trewin, Hazel R. Parry, Daniel E. Pagendam, Gregor J. Devine, Myron P. Zalucki, Jonathan M. Darbro, Cassie C. Jansen, and Nancy A. Schellhorn. Simulating an invasion: unsealed water storage (rainwater tanks) and urban block design facilitate the spread of the dengue fever mosquito, Aedes aegypti, in Brisbane, Australia. *Biological Invasions*, 23(12):3891–3906, December 2021. ISSN 1387-3547, 1573-1464. doi: 10.1007/s10530-021-02619-z. URL https://link.springer.com/10.1007/s10530-021-02619-z.

[528] Michael Kearney, Warren P. Porter, Craig Williams, Scott Ritchie, and Ary A. Hoffmann. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Functional Ecology*, 23(3):528–538, June 2009. ISSN 02698463, 13652435. doi: 10.1111/j.1365-2435.2008.01538.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2435.2008.01538.x.

[529] Cameron Nosrat, Jonathan Altamirano, Assaf Anyamba, Jamie M. Caldwell, Richard Damoah, Francis Mutuku, Bryson Ndenga, and A. Desiree LaBeaud. Impact of recent climate extremes on mosquito-borne disease transmission

in Kenya. *PLOS Neglected Tropical Diseases*, 15(3):e0009182, March 2021. ISSN 1935-2735. doi: 10.1371/journal.pntd.0009182. URL https://dx.plos.org/10.1371/journal.pntd.0009182.

[530] Simon Hales, Phil Weinstein, Yvan Souares, and Alistair Woodward. El Nino and the Dynamics of Vectorborne Disease Transmission. *Environmental Health Perspectives*, 107(2):99, February 1999. ISSN 00916765. doi: 10.2307/3434364. URL https://www.jstor.org/stable/3434364?origin=crossref.

[531] Tia Dostal, Julianne Meisner, César Munayco, Patricia J. García, César Cárcamo, Jose Enrique Pérez Lu, Cory Morin, Lauren Frisbie, and Peter M. Rabinowitz. The effect of weather and climate on dengue outbreak risk in Peru, 2000-2018: A time-series analysis. *PLOS Neglected Tropical Diseases*, 16(6):e0010479, June 2022. ISSN 1935-2735. doi: 10.1371/journal.pntd.0010479. URL https://dx.plos.org/10.1371/journal.pntd.0010479.

[532] Red Cross and Red Crescent Societies. Emergency Appeal Final Report. Peru: Floods. Technical report, December 2018. URL https://reliefweb.int/sites/reliefweb.int/files/resources/MDRPE012.efr_.pdf.

[533] Andréa Sobral De Almeida, Roberto De Andrade Medronho, and Luís Iván Ortiz Valencia. Análise espacial da dengue e o contexto socioeconômico no município do Rio de Janeiro, RJ. *Revista de Saúde Pública*, 43(4):666–673, August 2009. ISSN 0034-8910. doi: 10.1590/S0034-89102009000400013. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102009000400013&lng=pt&tlng=pt.

[534] Mary H. Hayden, Christopher K. Uejio, Kathleen Walker, Frank Ramberg, Rafael Moreno, Cecilia Rosales, Mercedes Gameros, Linda O. Mearns, Emily Zielinski-Gutierrez, and Craig R. Janes. Microclimate and Human Factors in the Divergent Ecology of Aedes aegypti along the Arizona, U.S./Sonora, MX Border. *EcoHealth*, 7(1):64–77, August 2010. ISSN 1612-9202, 1612-9210. doi: 10.1007/s10393-010-0288-z. URL http://link.springer.com/10.1007/s10393-010-0288-z.

[535] Jared Aldstadt, Constantianus J. M. Koenraadt, Thanyalak Fansiri, Udom Kijchalao, Jason Richardson, James W. Jones, and Thomas W. Scott. Ecological Modeling of Aedes aegypti (L.) Pupal Production in Rural Kamphaeng Phet, Thailand. *PLoS Neglected Tropical Diseases*, 5(1):e940, January 2011. ISSN 1935-2735. doi: 10.1371/journal.pntd.0000940. URL https://dx.plos.org/10.1371/journal.pntd.0000940.

[536] R J Pontes, A Spielman, J W Oliveira-Lima, J C Hodgson, and J Freeman. Vector densities that potentiate dengue outbreaks in a Brazilian city. *The American Journal of Tropical Medicine and Hygiene*, 62(3):378–383, March 2000. ISSN 0002-9637, 1476-1645. doi: 10.4269/ajtmh.2000.62.378. URL https://ajtmh.org/doi/10.4269/ajtmh.2000.62.378.

[537] Yoshiro Nagao, Usavadee Thavara, Pensri Chitnumsup, Apiwat Tawatsin, Chitti Chansang, and Diarmid Campbell-Lendrum. Climatic and social risk factors for Aedes infestation in rural Thailand. *Tropical Medicine and International Health*, 8(7):650–659, July 2003. ISSN 1360-2276, 1365-3156. doi: 10.1046/j.1365-3156.2003.01075.x. URL http://doi.wiley.com/10.1046/j.1365-3156.2003.01075.x.

[538] Leah C. Katzelnick, Lionel Gresh, M. Elizabeth Halloran, Juan Carlos Mercado, Guillermina Kuan, Aubree Gordon, Angel Balmaseda, and Eva Harris. Antibody-dependent enhancement of severe dengue disease in humans. *Science*, 358(6365):929–932, November 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aan6836. URL https://www.science.org/doi/10.1126/science.aan6836.

[539] Willem G. Van Panhuis, Marc Choisy, Xin Xiong, Nian Shong Chok, Pasakorn Akarasewi, Sopon Iamsirithaworn, Sai K. Lam, Chee K. Chong, Fook C. Lam, Bounlay Phommasak, Phengta Vongphrachanh, Khamphaphongphane

Bouaphanh, Huy Rekol, Nguyen Tran Hien, Pham Quang Thai, Tran Nhu Duong, Jen-Hsiang Chuang, Yu-Lun Liu, Lee-Ching Ng, Yuan Shi, Enrique A. Tayag, Vito G. Roque, Lyndon L. Lee Suy, Richard G. Jarman, Robert V. Gibbons, John Mark S. Velasco, In-Kyu Yoon, Donald S. Burke, and Derek A. T. Cummings. Region-wide synchrony and traveling waves of dengue across eight countries in Southeast Asia. *Proceedings of the National Academy of Sciences*, 112(42):13069–13074, October 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1501375112. URL https://pnas.org/doi/full/10.1073/pnas.1501375112.

[540] C. Codeco, F. Coelho, O. Cruz, S. Oliveira, T. Castro, and L. Bastos. Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. *Revue d'Épidémiologie et de Santé Publique*, 66:S386, July 2018. ISSN 03987620. doi: 10.1016/j.respe.2018.05.408. URL https://linkinghub.elsevier.com/retrieve/pii/S0398762018311088.

[541] Mazni Baharom, Norfazilah Ahmad, Rozita Hod, and Mohd Rizal Abdul Manaf. Dengue Early Warning System as Outbreak Prediction Tool: A Systematic Review. *Risk Management and Healthcare Policy*, Volume 15:871–886, May 2022. ISSN 1179-1594. doi: 10.2147/RMHP.S361106. URL https://www.dovepress.com/dengue-early-warning-system-as-outbreak-prediction-tool-a-systematic-r-peer-reviewed-fulltext-article-RMHP.

[542] Gina Tsarouchi and Darren Lumbroso. D-MOSS: An integrated dengue early warning system in Vietnam driven by Earth Observations. page 6962, April 2019. URL https://ui.adsabs.harvard.edu/abs/2019EGUGA..21.6962T. Conference Name: EGU General Assembly Conference Abstracts ADS Bibcode: 2019EGUGA..21.6962T.

[543] Mikhail Churakov, Christian J. Villabona-Arenas, Moritz U. G. Kraemer, Henrik Salje, and Simon Cauchemez. Spatio-temporal dynamics of dengue in Brazil: Seasonal travelling waves and determinants of regional synchrony. *PLOS Neglected Tropical Diseases*, 13(4):e0007012, April 2019. ISSN 1935-2735. doi: 10.1371/journal.pntd.0007012. URL https://dx.plos.org/10.1371/journal.pntd.0007012.

[544] G. Chowell, M. A. Miller, and C. Viboud. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiology and Infection*, 136(6):852–864, June 2008. ISSN 0950-2688, 1469-4409. doi: 10.1017/S0950268807009144. URL https://www.cambridge.org/core/product/identifier/S0950268807009144/type/journal_article.

[545] Léo Grinsztajn, Elizaveta Semenova, Charles C. Margossian, and Julien Riou. Bayesian workflow for disease transmission modeling in Stan. *Statistics in Medicine*, 40(27):6209–6234, November 2021. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.9164. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.9164.

[546] Z. F. Dembek, T. Chekol, and A. Wu. Best practice assessment of disease modelling for infectious disease outbreaks. *Epidemiology and Infection*, 146(10):1207–1215, 2018. doi: 10.1017/S095026881800119X.

[547] INEI. Censos Nacionales 2007: XI de Población y VI de Vivienda, 2007. URL https://www.gob.pe/36495-consultar-informacion-de-los-censos-realizados-por-el-inei-censos-nacionales-de-poblacion-y-vivienda.

[548] INEI. Resultados Definitivos de los Censos Nacionales 2017, 2017. URL https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1544/.

[549] INEI. Population and Housing, June 2022. URL https://www.inei.gob.pe/estadisticas/indice-tematico/poblacion-y-vivienda/.

[550] Angi Roesch and Harald Schmidbauer. *WaveletComp: Computational Wavelet Analysis*. 2018. URL https://CRAN.R-project.org/package=WaveletComp.

[551] Trevor Hastie and Robert Tibshirani. *Generalized additive models*. Chapman & Hall/CRC, Boca Raton, Fla, 1999. ISBN 978-0-412-34390-2.

[552] Simon N. Wood. The R Journal: Mgcv: GAMs and generalized ridge regression for R. *R News*, 1(2):20–25, 2001. ISSN 1609-3631.

[553] Peter Craven and Grace Wahba. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31(4):377–403, December 1978. ISSN 0029-599X, 0945-3245. doi: 10.1007/BF01404567. URL http://link.springer.com/10.1007/BF01404567.

[554] Roberto Buizza. Introduction to the special issue on "25 years of ensemble forecasting". *Quarterly Journal of the Royal Meteorological Society*, 145(S1):1–11, September 2019. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.3370. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3370.

[555] Madhurima Panja, Tanujit Chakraborty, Sk Shahid Nadim, Indrajit Ghosh, Uttam Kumar, and Nan Liu. An ensemble neural network approach to forecast Dengue outbreak based on climatic condition. *Chaos, Solitons & Fractals*, 167: 113124, February 2023. ISSN 09600779. doi: 10.1016/j.chaos.2023.113124. URL https://linkinghub.elsevier.com/retrieve/pii/S0960077923000255.

[556] Soudeep Deb and Sougata Deb. An Ensemble Method for Early Prediction of Dengue Outbreak. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(1):84–101, January 2022. ISSN 0964-1998, 1467-985X. doi: 10.1111/rssa.12714. URL https://academic.oup.com/jrsssa/article/185/1/84/7068440.

[557] Alessandro Sebastianelli, Dario Spiller, Raquel Carmo, James Wheeler, Artur Nowakowski, Ludmilla Viana Jacobson, Dohyung Kim, Hanoch Barlevi, Zoraya El Raiss Cordero, Felipe J Colón-González, Rachel Lowe, Silvia Liberata Ullo, and Rochelle Schneider. A reproducible ensemble machine learning approach to forecast dengue outbreaks. *Scientific Reports*, 14(1):3807, February 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-52796-9. URL https://www.nature.com/articles/s41598-024-52796-9.

[558] Ryan Tibshirani and Logan Brooks. *quantgen: Tools for generalized quantile modeling*. 2022.

[559] Rob J. Hyndman. Forecasting, causality and feedback. *International Journal of Forecasting*, 39(2):558–560, April 2023. ISSN 01692070. doi: 10.1016/j.ijforecast.2022.09.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0169207022001388.

[560] Nikos I. Bosse, Hugo Gruson, Anne Cori, Edwin van Leeuwen, Sebastian Funk, and Sam Abbott. Evaluating Forecasts with scoringutils in R, November 2024. URL http://arxiv.org/abs/2205.07090. arXiv:2205.07090 [stat].

[561] Julien Herzen, Francesco LÃ¤ssig, Samuele Giuliano Piazzetta, Thomas Neuer, Leo Tafti, Guillaume Raille, Tomas Van Pottelbergh, Marek Pasieka, Andrzej Skrodzki, Nicolas Huguenin, Maxime Dumonal, Jan KoÅ›cisz, Dennis Bader, FrÃ©dÃ©rick Gusset, Mounir Benheddi, Camila Williamson, Michal Kosinski, Matej Petrik, and GaÃ«l Grosch. Darts: User-Friendly Modern Machine Learning for Time Series. *Journal of Machine Learning Research*, 23(124):1–6, 2022. URL http://jmlr.org/papers/v23/21-1177.html.

[562] Cristian Challú Kin G. Olivares Federico Garza, Max Mergenthaler Canseco. StatsForecast: Lightning fast forecasting with statistical and econometric models, 2022. URL https://github.com/Nixtla/statsforecast. Published: PyCon Salt Lake City, Utah, US 2022.

[563] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, August 2023. URL http://arxiv.org/abs/1706.03762. arXiv:1706.03762 [cs].

[564] Simon Pollett, Michael A. Johansson, Nicholas G. Reich, David Brett-Major, Sara Y. Del Valle, Srinivasan Venkatramanan, Rachel Lowe, Travis Porco, Irina Maljkovic Berry, Alina Deshpande, Moritz U. G. Kraemer, David L. Blazes, Wirichada Pan-ngum, Alessandro Vespigiani, Suzanne E. Mate, Sheetal P. Silal, Sasikiran Kandula, Rachel Sippy, Talia M. Quandelacy, Jeffrey J. Morgan, Jacob Ball, Lindsay C. Morton, Benjamin M. Althouse, Julie Pavlin, Wilbert Van Panhuis, Steven Riley, Matthew Biggerstaff, Cecile Viboud, Oliver Brady, and Caitlin Rivers. Recommended reporting items for epidemic forecasting and prediction research: The EPIFORGE 2020 guidelines. *PLOS Medicine*, 18(10):e1003793, October 2021. ISSN 1549-1676. doi: 10.1371/journal.pmed.1003793. URL https://dx.plos.org/10.1371/journal.pmed.1003793.

[565] Gene H. Golub, Michael Heath, and Grace Wahba. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21(2):215–223, May 1979. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.1979.10489751. URL http://www.tandfonline.com/doi/abs/10.1080/00401706.1979.10489751.

[566] Clifford M. Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/76.2.297. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/76.2.297.

[567] C. MacLachlan, A. Arribas, K. A. Peterson, A. Maidens, D. Fereday, A. A. Scaife, M. Gordon, M. Vellinga, A. Williams, R. E. Comer, J. Camp, P. Xavier, and G. Madec. Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1072–1084, April 2015. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.2396. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.2396.

[568] Osama M. E. Seidahmed and Elfatih A. B. Eltahir. A Sequence of Flushing and Drying of Breeding Habitats of Aedes aegypti (L.) Prior to the Low Dengue Season in Singapore. *PLOS Neglected Tropical Diseases*, 10(7):e0004842, July 2016. ISSN 1935-2735. doi: 10.1371/journal.pntd.0004842. URL https://dx.plos.org/10.1371/journal.pntd.0004842.

[569] World Health Organization and UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. *Global vector control response 2017-2030*. World Health Organization, Geneva, 2017. ISBN 978-92-4-151297-8. URL https://iris.who.int/handle/10665/259205.

[570] Taha Falatouri, Farzaneh Darbanian, Patrick Brandtner, and Chibuzor Udokwu. Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. *Procedia Computer Science*, 200:993–1003, 2022. ISSN 18770509. doi: 10.1016/j.procs.2022.01.298. URL https://linkinghub.elsevier.com/retrieve/pii/S1877050922003076.

[571] Lorenzo Menculini, Andrea Marini, Massimiliano Proietti, Alberto Garinei, Alessio Bozza, Cecilia Moretti, and Marcello Marconi. Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. *Forecasting*, 3(3):644–662, September 2021. ISSN 2571-9394. doi: 10.3390/forecast3030040. URL https://www.mdpi.com/2571-9394/3/3/40.

[572] Tim K. Tsang, Qiurui Du, Benjamin J. Cowling, and Cécile Viboud. An adaptive weight ensemble approach to forecast influenza activity in an irregular seasonality context. *Nature Communications*, 15(1):8625, October 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-52504-1. URL https://www.nature.com/articles/s41467-024-52504-1.

[573] Kelly Charniga, Zachary J. Madewell, Nina B. Masters, Jason Asher, Yoshinori Nakazawa, and Ian H. Spicknall. Nowcasting and forecasting the 2022 U.S. mpox outbreak: Support for public health decision making and lessons learned. *Epidemics*, 47:100755, June 2024. ISSN 1755-4365. doi: 10.1016/j.epidem.2024.100755. URL https://www.sciencedirect.com/science/article/pii/S1755436524000161.

[574] Emilie Finch, Adam Kucharski, Shuzhen Sim, Lee-Ching Ng, and Rachel Lowe. Climate variation and serotype competition drive dengue outbreak dynamics in Singapore, September 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.09.17.24313793.

[575] Guanghu Zhu, Tao Liu, Jianpeng Xiao, Bing Zhang, Tie Song, Yonghui Zhang, Lifeng Lin, Zhiqiang Peng, Aiping Deng, Wenjun Ma, and Yuantao Hao. Effects of human mobility, temperature and mosquito control on the spatiotemporal transmission of dengue. *Science of The Total Environment*, 651:969–978, February 2019. ISSN 00489697. doi: 10.1016/j.scitotenv.2018.09.182. URL https://linkinghub.elsevier.com/retrieve/pii/S004896971833626X.

[576] Emanuele Massaro, Daniel Kondor, and Carlo Ratti. Assessing the interplay between human mobility and mosquito borne diseases in urban environments. *Scientific Reports*, 9(1):16911, November 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-53127-z. URL https://www.nature.com/articles/s41598-019-53127-z.

[577] Mathew V. Kiang, Mauricio Santillana, Jarvis T. Chen, Jukka-Pekka Onnela, Nancy Krieger, Kenth Engø-Monsen, Nattwut Ekapirat, Darin Areechokchai, Preecha Prempree, Richard J. Maude, and Caroline O. Buckee. Incorporating human mobility data improves forecasts of Dengue fever in Thailand. *Scientific Reports*, 11(1):923, January 2021. ISSN 2045-2322. doi: 10.1038/s41598-020-79438-0. URL https://www.nature.com/articles/s41598-020-79438-0.

[578] Brazilian Institute of Geography and Statistics. 2022 Brazilian Census, 2022. URL https://censo2022.ibge.gov.br/panorama/.

[579] Benjamin MacCormack-Gelles, Antonio S. Lima Neto, Geziel S. Sousa, Osmar J. Nascimento, Marcia M. T. Machado, Mary E. Wilson, and Marcia C. Castro. Epidemiological characteristics and determinants of dengue transmission during epidemic and non-epidemic years in Fortaleza, Brazil: 2011-2015. *PLOS Neglected Tropical Diseases*, 12(12):e0006990, December 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006990. URL https://dx.plos.org/10.1371/journal.pntd.0006990.

[580] Daniela Cristina Sensato Monteiro, Natália Vasconcelos De Souza, Jane Cavalcante Amaral, Kaynan Bezerra De Lima, Fernanda Montenegro Carvalho De Araújo, Izabel Letícia Cavalcante Ramalho, Victor Emanuel Pessoa Martins, Jeová Keny Baima Colares, Luciano Pamplona De Góes Cavalcanti, and Danielle Malta Lima. Dengue: 30 years of cases in an endemic area. *Clinics*, 74:e675, 2019. ISSN 18075932. doi: 10.6061/clinics/2019/e675. URL https://linkinghub.elsevier.com/retrieve/pii/S1807593222006342.

[581] Carlos Caminha, Vasco Furtado, Vladia Pinheiro, and Caio Silva. Micro-interventions in urban transportation from pattern discovery on the flow of passengers and on the bus network. In *2016 IEEE International Smart Cities Conference (ISC2)*, pages 1–6, Trento, Italy, September 2016. IEEE. ISBN 978-1-5090-1846-8. doi: 10.1109/ISC2.2016.7580776. URL http://ieeexplore.ieee.org/document/7580776/.

[582] Carlos Caminha and Vasco Furtado. Impact of human mobility on police allocation. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 125–127, Beijing, China, July 2017. IEEE. ISBN 978-1-5090-6727-5. doi: 10.1109/ISI.2017.8004886. URL http://ieeexplore.ieee.org/document/8004886/.

[583] Carlos Caminha, Vasco Furtado, Tarcisio H. C. Pequeno, Caio Ponte, Hygor P. M. Melo, Erneson A. Oliveira, and José S. Andrade. Human mobility in large cities as a proxy for crime. *PLOS ONE*, 12(2):e0171609, February 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0171609. URL https://dx.plos.org/10.1371/journal.pone.0171609.

[584] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL https://www.R-project.org/.

[585] Jonah Gabry, Rok Češnovar, Andrew Johnson, and Steve Bronder. *cmdstanr: R Interface to 'CmdStan'*. 2024. URL https://mc-stan.org/cmdstanr/.

[586] Karl Pearson. Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London Series I*, 58:240–242, January 1895. URL https://ui.adsabs.harvard.edu/abs/1895RSPS...58..240P. ADS Bibcode: 1895RSPS...58..240P.

[587] C. Spearman. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1):72, January 1904. ISSN 00029556. doi: 10.2307/1412159. URL https://www.jstor.org/stable/1412159?origin=crossref.

[588] Olive Jean Dunn. Multiple Comparisons among Means. *Journal of the American Statistical Association*, 56 (293):52–64, March 1961. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1961.10482090. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1961.10482090.

[589] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008. ISSN 1742-5468. doi: 10.1088/1742-5468/2008/10/P10008. URL http://arxiv.org/abs/0803.0476. arXiv:0803.0476 [physics].

[590] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, January 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0706851105. URL http://arxiv.org/abs/0707.0609. arXiv:0707.0609 [physics].

[591] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948. ISSN 00058580. doi: 10.1002/j.1538-7305.1948.tb01338.x. URL https://ieeexplore.ieee.org/document/6773024.

[592] Eduardo Correa Araujo, Luiz Max Carvalho, Fabiana Ganem, Luã Bida Vacaro, Leonardo S. Bastos, Laís Picinini Freitas, Marcio Bastos, Ramila Alencar, Lucas Bianchi, Raúl Capellán, Xiang Chen, Oswaldo Cruz, Americo Cunha, Haridas K. Das, Chloe Fletcher, Raquel Martins Lana, Rachel Lowe, Daniela Lührsen, Giovenale Moirano, Paula Moraga, Lucas M. Stolerman, Fernanda Valente, Cláudia Torres Codeço, and Flávio C. Coelho. Leveraging probabilistic forecasts for dengue preparedness and control: the 2024 Dengue Forecasting Sprint in Brazil, May 2025. URL http://medrxiv.org/lookup/doi/10.1101/2025.05.12.25327419.

[593] Thales Bermann, Tatiana Schaffer Gregianini, Fernanda Marques Godinho, Ludmila Fiorenzano Baethgen, Fernanda Letícia Martiny, Erica Bortoli Möllmann, Amanda Pellenz Ruivo, Franciellen Machado dos Santos, Caroline do Nascimento Ferreira, Gustavo Barbosa de Lima, Danielly Alves Mendes Barbosa, Filipe Zimmer Dezordi, Milena Bauermann, Taina Machado Selayaran, Elgion Lucio da Silva Loreto, Felipe Gomes Naveca, André Luiz Sá de Oliveira, Ana Beatriz Gorini da Veiga, Marcelo Henrique Santos Paiva, de Lima Tulio Campos, Gonzalo Bello, Matheus Filgueira Bezerra, Felipe Valter de Oliveira, Rafael Maciel-de Freitas, Tiago Gräf,

Richard Steiner Salvato, and Gabriel Luz Wallau. Increasing dengue outbreaks in temperate Brazil is linked to Aedes aegypti invasion and infestation level driving widespread virus transmission, February 2025. URL https://www.medrxiv.org/content/10.1101/2025.02.24.25322773v1. Pages: 2025.02.24.25322773.

[594] A. J. Tatem. Mapping population and pathogen movements. *International Health*, 6(1):5–11, March 2014. ISSN 1876-3413, 1876-3405. doi: 10.1093/inthealth/ihu006. URL https://academic.oup.com/inthealth/article-lookup/doi/10.1093/inthealth/ihu006.

[595] Oriol Cabanas-Tirapu, Lluís Danús, Esteban Moro, Marta Sales-Pardo, and Roger Guimerà. Human mobility is well described by closed-form gravity-like models learned automatically from data. *Nature Communications*, 16(1):1336, February 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-56495-5. URL https://www.nature.com/articles/s41467-025-56495-5.

[596] Javier Aguilar, Aleix Bassolas, Gourab Ghoshal, Surendra Hazarie, Alec Kirkley, Mattia Mazzoli, Sandro Meloni, Sayat Mimar, Vincenzo Nicosia, José J. Ramasco, and Adam Sadilek. Impact of urban structure on infectious disease spreading. *Scientific Reports*, 12(1):3816, March 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-06720-8.

[597] Ake Svensson. A note on generation times in epidemic models. *Mathematical Biosciences*, 208(1):300–311, July 2007. ISSN 00255564. doi: 10.1016/j.mbs.2006.10.010. URL https://linkinghub.elsevier.com/retrieve/pii/S0025556406002094.

[598] I Ogi-Gittins, Ws Hart, J Song, Rk Nash, J Polonsky, A Cori, Em Hill, and Rn Thompson. A simulation-based approach for estimating the time-dependent reproduction number from temporally aggregated disease incidence time series data, September 2023. URL http://medrxiv.org/lookup/doi/10.1101/2023.09.13.23295471.

[599] David Champredon, Jonathan Dushoff, and David J. D. Earn. Equivalence of the Erlang-Distributed SEIR Epidemic Model and the Renewal Equation. *SIAM Journal on Applied Mathematics*, 78(6):3258–3278, January 2018. ISSN 0036-1399, 1095-712X. doi: 10.1137/18M1186411. URL https://epubs.siam.org/doi/10.1137/18M1186411.

[600] Hisashi Inaba and Hiroshi Nishiura. The state-reproduction number for a multistate class age structured epidemic system and its application to the asymptomatic transmission model. *Mathematical Biosciences*, 216(1):77–89, November 2008. ISSN 00255564. doi: 10.1016/j.mbs.2008.08.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0025556408001296.

[601] Victoria Romeo-Aznar, Olivier Telle, Mauricio Santos-Vega, Richard Paul, and Mercedes Pascual. Crowded and warmer: Unequal dengue risk at high spatial resolution across a megacity of India. *PLOS Climate*, 3(3):e0000240, March 2024. ISSN 2767-3200. doi: 10.1371/journal.pclm.0000240. URL https://dx.plos.org/10.1371/journal.pclm.0000240.

[602] T. Alex Perkins, C. Jessica E. Metcalf, Bryan T. Grenfell, and Andrew J. Tatem. Estimating drivers of autochthonous transmission of chikungunya virus in its invasion of the americas. *PLoS currents*, 7: ecurrents.outbreaks.a4c7b6ac10e0420b1788c9767946d1fc, February 2015. ISSN 2157-3999. doi: 10.1371/currents.outbreaks.a4c7b6ac10e0420b1788c9767946d1fc.

[603] James Dickson Murray. *Mathematical biology*. Number 17-18 in Interdisciplinary applied mathematics. Springer, New York [etc.], 3rd ed edition, 2002. ISBN 978-0-387-95223-9 978-0-387-95228-4.

[604] Lorenzo Pellis, Paul J. Birrell, Joshua Blake, Christopher E. Overton, Francesca Scarabel, Helena B. Stage, Ellen Brooks-Pollock, Leon Danon, Ian Hall, Thomas A. House, Matt J. Keeling, Jonathan M. Read, JUNIPER Consortium,

and Daniela De Angelis. Estimation of Reproduction Numbers in Real Time: Conceptual and Statistical Challenges. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(Supplement_1):S112–S130, November 2022. ISSN 0964-1998. doi: 10.1111/rssa.12955. URL https://doi.org/10.1111/rssa.12955.

[605] Johan A. J. Metz and Odo Diekmann. *The dynamics of physiologically structured populations*. Number 68 in Lecture notes in biomathematics. Springer-Verl, Berlin, 1986. ISBN 978-0-387-16786-2 978-3-540-16786-0.

[606] D. Breda, O. Diekmann, W. F. De Graaf, A. Pugliese, and R. Vermiglio. On the formulation of epidemic models (an appraisal of Kermack and McKendrick). *Journal of Biological Dynamics*, 6:103–117, September 2012. ISSN 1751-3758, 1751-3766. doi: 10.1080/17513758.2012.716454. URL http://www.tandfonline.com/doi/abs/10.1080/17513758.2012.716454.

[607] Sebastian Funk, Adam J. Kucharski, Anton Camacho, Rosalind M. Eggo, Laith Yakob, Lawrence M. Murray, and W. John Edmunds. Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. *PLOS Neglected Tropical Diseases*, 10(12):e0005173, December 2016. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005173. URL https://dx.plos.org/10.1371/journal.pntd.0005173.

[608] Ben Lambert, Ace North, and H. Charles J. Godfray. A meta-analysis of longevity estimates of mosquito Vectors of disease, May 2022. URL http://biorxiv.org/lookup/doi/10.1101/2022.05.30.494059.

[609] Meteorological Service Singapore. Historical Daily Records, August 2024. URL https://www.weather.gov.sg/climate-historical-daily/.

[610] Cathal Mills. Repository for renewal equations for mosquito-borne diseases, June 2025. URL https://github.com/cathalmills/vbd_renewal.

[611] Marta S Shocket, Anna B Verwillow, Mailo G Numazu, Hani Slamani, Jeremy M Cohen, Fadoua El Moustaid, Jason Rohr, Leah R Johnson, and Erin A Mordecai. Transmission of West Nile and five other temperate mosquito-borne viruses peaks at temperatures between 23°C and 26°C. *eLife*, 9:e58511, September 2020. ISSN 2050-084X. doi: 10.7554/eLife.58511. URL https://elifesciences.org/articles/58511.

[612] Blanka Tesla, Leah R. Demakovsky, Erin A. Mordecai, Sadie J. Ryan, Matthew H. Bonds, Calistus N. Ngonghala, Melinda A. Brindley, and Courtney C. Murdock. Temperature drives Zika virus transmission: evidence from empirical and mathematical models. *Proceedings of the Royal Society B: Biological Sciences*, 285(1884):20180795, August 2018. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2018.0795. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2018.0795.

[613] William Michael Landau. The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, 6(57):2959, 2021. URL https://doi.org/10.21105/joss.02959.

[614] Radek Erban, Jonathan Chapman, and Philip Maini. A practical guide to stochastic simulations of reaction-diffusion processes, November 2007. URL http://arxiv.org/abs/0704.1908. arXiv:0704.1908 [q-bio].

[615] Daniel T. Gillespie. Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry*, 58(Volume 58, 2007):35–55, May 2007. ISSN 0066-426X, 1545-1593. doi: 10.1146/annurev.physchem.58.032806.104637. URL https://www.annualreviews.org/content/journals/10.1146/annurev.physchem.58.032806.104637. Publisher: Annual Reviews.

[616] Desmond J. Higham. Modeling and Simulating Chemical Reactions. *SIAM Review*, 50(2):347–368, January 2008. ISSN 0036-1445. doi: 10.1137/060666457. URL https://epubs.siam.org/doi/10.1137/060666457. Publisher: Society for Industrial and Applied Mathematics.

[617] Adam Arkin, John Ross, and Harley H McAdams. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage -Infected Escherichia coli Cells. *Genetics*, 149(4):1633–1648, August 1998. ISSN 1943-2631. doi: 10.1093/genetics/149.4.1633. URL https://academic.oup.com/genetics/article/149/4/1633/6034258.

[618] H. Li, Y. Cao, L.R. Petzold, and D.T. Gillespie. Algorithms and Software for Stochastic Simulation of Biochemical Reacting Systems. *Biotechnology Progress*, 24(1):56–61, February 2008. ISSN 8756-7938. doi: 10.1021/bp070255h. URL http://doi.wiley.com/10.1021/bp070255h.

[619] Michael A. Gibson and Jehoshua Bruck. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, March 2000. ISSN 1089-5639, 1520-5215. doi: 10.1021/jp993732q. URL https://pubs.acs.org/doi/10.1021/jp993732q.

[620] Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, July 2001. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.1378322. URL https://pubs.aip.org/jcp/article/115/4/1716/451187/Approximate-accelerated-stochastic-simulation-of.

[621] Yang Cao, Daniel T. Gillespie, and Linda R. Petzold. Avoiding negative populations in explicit Poisson tau-leaping. *The Journal of Chemical Physics*, 123(5):054104, August 2005. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.1992473. URL https://pubs.aip.org/jcp/article/123/5/054104/905859/Avoiding-negative-populations-in-explicit-Poisson.

[622] Yang Cao, Daniel T. Gillespie, and Linda R. Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics*, 124(4):044109, January 2006. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.2159468. URL https://pubs.aip.org/jcp/article/124/4/044109/562210/Efficient-step-size-selection-for-the-tau-leaping.

[623] Alvaro Moraes, Raul Tempone, and Pedro Vilanova. Hybrid Chernoff Tau-Leap. *Multiscale Modeling & Simulation*, 12(2):581–615, January 2014. ISSN 1540-3459. doi: 10.1137/130925657. URL https://epubs.siam.org/doi/10.1137/130925657. Publisher: Society for Industrial and Applied Mathematics.

[624] David F. Anderson. A modified Next Reaction Method for simulating chemical systems with time dependent propensities and delays. *The Journal of Chemical Physics*, 127(21):214107, December 2007. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.2799998. URL http://arxiv.org/abs/0708.0370. arXiv:0708.0370 [q-bio].

[625] James P Roney, Jeremy Ferlic, Franziska Michor, and Thomas O McDonald. ESTIpop: a computational tool to simulate and estimate parameters for continuous-time Markov branching processes. *Bioinformatics*, 36(15):4372–4373, August 2020. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btaa526. URL https://academic.oup.com/bioinformatics/article/36/15/4372/5840725.

[626] P. A. W Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, September 1979. ISSN 0028-1441, 1931-9193. doi: 10.1002/nav.3800260304. URL https://onlinelibrary.wiley.com/doi/10.1002/nav.3800260304.

[627] Karline Soetaert, Thomas Petzoldt, and R. Woodrow Setzer. Solving Differential Equations in *R* : Package **deSolve**. *Journal of Statistical Software*, 33(9), 2010. ISSN 1548-7660. doi: 10.18637/jss.v033.i09. URL http://www.jstatsoft.org/v33/i09/.

[628] Linda Petzold. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM Journal on Scientific and Statistical Computing*, 4(1):136–148, March 1983. ISSN 0196-5204, 2168-3417. doi: 10.1137/0904010. URL http://epubs.siam.org/doi/10.1137/0904010.

[629] A Hindmarsh. ODEPACK. A Collection of ODE System Solvers. Technical report, United States, 2016. URL https://www.osti.gov/biblio/145724.

[630] Fabian Spill, Philip K. Maini, and Helen Byrne. Optimisation of Simulations of Stochastic Processes by Removal of Opposing Reactions. 2016. doi: 10.48550/ARXIV.1602.02655. URL https://arxiv.org/abs/1602.02655. Publisher: arXiv Version Number: 1.

[631] Nicholas G. Reich, Logan C. Brooks, Spencer J. Fox, Sasikiran Kandula, Craig J. McGowan, Evan Moore, Dave Osthus, Evan L. Ray, Abhinav Tushar, Teresa K. Yamana, Matthew Biggerstaff, Michael A. Johansson, Roni Rosenfeld, and Jeffrey Shaman. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences*, 116(8):3146–3154, February 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1812594116. URL https://pnas.org/doi/full/10.1073/pnas.1812594116.

[632] Nikos I. Bosse, Sam Abbott, Johannes Bracher, Habakuk Hain, Billy J. Quilty, Mark Jit, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Edwin Van Leeuwen, Anne Cori, and Sebastian Funk. Comparing human and model-based forecasts of COVID-19 in Germany and Poland. *PLOS Computational Biology*, 18(9):e1010405, September 2022. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1010405. URL https://dx.plos.org/10.1371/journal.pcbi.1010405.

[633] Jonathon Mellor, Maria L Tang, Owen Jones, Thomas Ward, Steven Riley, and Sarah R Deeny. Forecasting COVID-19, influenza, and RSV hospitalizations over winter 2023–4 in England. *International Journal of Epidemiology*, 54 (3):dyaf066, April 2025. ISSN 0300-5771, 1464-3685. doi: 10.1093/ije/dyaf066. URL https://academic.oup.com/ije/article/doi/10.1093/ije/dyaf066/8156944.

[634] Martin Janousek Thomas Haiden. Evaluation of ECMWF forecasts, including the 2023 upgrade, 2023. URL https://www.ecmwf.int/en/elibrary/81389-evaluation-ecmwf-forecasts-including-2023-upgrade.

[635] Uwe Siebert. When should decision-analytic modeling be used in the economic evaluation of health care? *The European Journal of Health Economics, formerly: HEPAC*, 4(3):143–150, September 2003. ISSN 1618-7601. doi: 10.1007/s10198-003-0205-2. URL https://doi.org/10.1007/s10198-003-0205-2.

[636] Vanessa J. Fundel, Nadine Fleischhut, Stefan M. Herzog, Martin Göber, and Renate Hagedorn. Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *Quarterly Journal of the Royal Meteorological Society*, 145(S1):210–231, September 2019. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.3482. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3482.

[637] Glenn W. Brier. Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78: 1, January 1950. ISSN 0027-0644. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2. URL https://ui.adsabs.harvard.edu/abs/1950MWRv...78....1B. Publisher: AMS ADS Bibcode: 1950MWRv...78....1B.

[638] Estee Y Cramer, Evan L Ray, Velma K Lopez, Johannes Bracher, Andrea Brennen, Alvaro J Castro Rivadeneira, Aaron Gerding, Tilmann Gneiting, Katie H House, Yuxin Huang, Dasuni Jayawardena, Abdul H Kanji, Ayush Khandelwal, Khoa Le, Anja Mühlemann, Jarad Niemi, Apurv Shah, Ariane Stark, Yijin Wang, Nutcha Wattanachit, Martha W Zorn, Youyang Gu, Sansiddh Jain, Nayana Bannur, Ayush Deva, Mihir Kulkarni, Srujana Merugu, Alpan

Raval, Siddhant Shingi, Avtansh Tiwari, Jerome White, Neil F Abernethy, Spencer Woody, Maytal Dahan, Spencer Fox, Kelly Gaither, Michael Lachmann, Lauren Ancel Meyers, James G Scott, Mauricio Tec, Ajitesh Srivastava, Glover E George, Jeffrey C Cegan, Ian D Dettwiller, William P England, Matthew W Farthing, Robert H Hunter, Brandon Lafferty, Igor Linkov, Michael L Mayo, Matthew D Parno, Michael A Rowland, Benjamin D Trump, Yanli Zhang-James, Samuel Chen, Stephen V Faraone, Jonathan Hess, Christopher P Morley, Asif Salekin, Dongliang Wang, Sabrina M Corsetti, Thomas M Baer, Marisa C Eisenberg, Karl Falb, Yitao Huang, Emily T Martin, Ella McCauley, Robert L Myers, Tom Schwarz, Daniel Sheldon, Graham Casey Gibson, Rose Yu, Liyao Gao, Yian Ma, Dongxia Wu, Xifeng Yan, Xiaoyong Jin, Yu-Xiang Wang, YangQuan Chen, Lihong Guo, Yanting Zhao, Quanquan Gu, Jinghui Chen, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Hannah Biegel, Joceline Lega, Steve McConnell, Vp Nagraj, Stephanie L Guertin, Christopher Hulme-Lowe, Stephen D Turner, Yunfeng Shi, Xuegang Ban, Robert Walraven, Qi-Jun Hong, Stanley Kong, Axel Van De Walle, James A Turtle, Michal Ben-Nun, Steven Riley, Pete Riley, Ugur Koyluoglu, David DesRoches, Pedro Forli, Bruce Hamory, Christina Kyriakides, Helen Leis, John Milliken, Michael Moloney, James Morgan, Ninad Nirgudkar, Gokce Ozcan, Noah Piwonka, Matt Ravi, Chris Schrader, Elizabeth Shakhnovich, Daniel Siegel, Ryan Spatz, Chris Stiefeling, Barrie Wilkinson, Alexander Wong, Sean Cavany, Guido España, Sean Moore, Rachel Oidtman, Alex Perkins, David Kraus, Andrea Kraus, Zhifeng Gao, Jiang Bian, Wei Cao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Alessandro Vespignani, Matteo Chinazzi, Jessica T Davis, Kunpeng Mu, Ana Pastore Y Piontti, Xinyue Xiong, Andrew Zheng, Jackie Baek, Vivek Farias, Andreea Georgescu, Retsef Levi, Deeksha Sinha, Joshua Wilde, Georgia Perakis, Mohammed Amine Bennouna, David Nze-Ndong, Divya Singhvi, Ioannis Spantidakis, Leann Thayaparan, Asterios Tsiourvas, Arnab Sarker, Ali Jadbabaie, Devavrat Shah, Nicolas Della Penna, Leo A Celi, Saketh Sundar, Russ Wolfinger, Dave Osthus, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dean Karlen, Matt Kinsey, Luke C. Mullany, Kaitlin Rainwater-Lovett, Lauren Shin, Katharine Tallaksen, Shelby Wilson, Elizabeth C Lee, Juan Dent, Kyra H Grantz, Alison L Hill, Joshua Kaminsky, Kathryn Kaminsky, Lindsay T Keegan, Stephen A Lauer, Joseph C Lemaitre, Justin Lessler, Hannah R Meredith, Javier Perez-Saez, Sam Shah, Claire P Smith, Shaun A Truelove, Josh Wills, Maximilian Marshall, Lauren Gardner, Kristen Nixon, John C. Burant, Lily Wang, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Guannan Wang, Yueying Wang, Shan Yu, Robert C Reiner, Ryan Barber, Emmanuela Gakidou, Simon I. Hay, Steve Lim, Chris J.L. Murray, David Pigott, Heidi L Gurung, Prasith Baccam, Steven A Stage, Bradley T Suchoski, B. Aditya Prakash, Bijaya Adhikari, Jiaming Cui, Alexander Rodríguez, Anika Tabassum, Jiajia Xie, Pinar Keskinocak, John Asplund, Arden Baxter, Buse Eylul Oruc, Nicoleta Serban, Sercan O Arik, Mike Dusenberry, Arkady Epshteyn, Elli Kanal, Long T Le, Chun-Liang Li, Tomas Pfister, Dario Sava, Rajarishi Sinha, Thomas Tsai, Nate Yoder, Jinsung Yoon, Leyou Zhang, Sam Abbott, Nikos I Bosse, Sebastian Funk, Joel Hellewell, Sophie R Meakin, Katharine Sherratt, Mingyuan Zhou, Rahi Kalantari, Teresa K Yamana, Sen Pei, Jeffrey Shaman, Michael L Li, Dimitris Bertsimas, Omar Skali Lami, Saksham Soni, Hamza Tazi Bouardi, Turgay Ayer, Madeline Adee, Jagpreet Chhatwal, Ozden O Dalgic, Mary A Ladd, Benjamin P Linas, Peter Mueller, Jade Xiao, Yuanjia Wang, Qinxia Wang, Shanghong Xie, Donglin Zeng, Alden Green, Jacob Bien, Logan Brooks, Addison J Hu, Maria Jahja, Daniel McDonald, Balasubramanian Narasimhan, Collin Politsch, Samyak Rajanala, Aaron Rumack, Noah Simon, Ryan J Tibshirani, Rob Tibshirani, Valerie Ventura, Larry Wasserman, Eamon B O'Dea, John M Drake, Robert Pagano, Quoc T Tran, Lam Si Tung Ho, Huong Huynh, Jo W Walker, Rachel B Slayton, Michael A Johansson, Matthew Biggerstaff, and Nicholas G Reich. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US, February 2021. URL `http://medrxiv.org/lookup/doi/10.1101/2021.02.03.21250974`.

[639] Edward Wheatcroft. Interpreting the skill score form of forecast performance metrics. *International Journal of Forecasting*, 35(2):573–579, April 2019. ISSN 01692070. doi: 10.1016/j.ijforecast.2018.11.010. URL `https://linkinghub.elsevier.com/retrieve/pii/S0169207019300093`.

[640] David Richardson. Skill and relative economic value of the ECMWF Ensemble Prediction System. 1998. doi: 10.21957/2XJS554NR. URL https://www.ecmwf.int/node/11903. Publisher: ECMWF.

[641] Timo Dimitriadis, Tilmann Gneiting, and Alexander I. Jordan. Stable reliability diagrams for probabilistic classifiers. *Proceedings of the National Academy of Sciences*, 118(8):e2016191118, February 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2016191118. URL https://pnas.org/doi/full/10.1073/pnas.2016191118.

[642] Allan H. Murphy. A New Vector Partition of the Probability Score. *Journal of Applied Meteorology*, 12(4): 595–600, June 1973. ISSN 0021-8952. doi: 10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2. URL http://journals.ametsoc.org/doi/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.

[643] Hans Hersbach. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, 15(5):559–570, October 2000. ISSN 0882-8156, 1520-0434. doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2. URL http://journals.ametsoc.org/doi/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

[644] G. Candille and O. Talagrand. Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society*, 131(609):2131–2150, July 2005. ISSN 0035-9009, 1477-870X. doi: 10.1256/qj.04.71. URL https://rmets.onlinelibrary.wiley.com/doi/10.1256/qj.04.71.

[645] Meelis Kull and Peter Flach. Novel Decompositions of Proper Scoring Rules for Classification: Score Adjustment as Precursor to Calibration. In Annalisa Appice, Pedro Pereira Rodrigues, Vítor Santos Costa, Carlos Soares, João Gama, and Alípio Jorge, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 9284, pages 68–85. Springer International Publishing, Cham, 2015. ISBN 978-3-319-23527-1 978-3-319-23528-8. doi: 10.1007/978-3-319-23528-8_5. URL http://link.springer.com/10.1007/978-3-319-23528-8_5. Series Title: Lecture Notes in Computer Science.

[646] Stefan Siegert. Simplifying and generalising Murphy's Brier score decomposition. *Quarterly Journal of the Royal Meteorological Society*, 143(703):1178–1183, January 2017. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.2985. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.2985.

[647] Sebastian Arnold, Eva-Maria Walz, Johanna Ziegel, and Tilmann Gneiting. Decompositions of the mean continuous ranked probability score. *Electronic Journal of Statistics*, 18(2), January 2024. ISSN 1935-7524. doi: 10.1214/24-EJS2316. URL https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-18/issue-2/Decompositions-of-the-mean-continuous-ranked-probability-score/10.1214/24-EJS2316.full.

[648] Tilmann Gneiting, Daniel Wolffram, Johannes Resin, Kristof Kraus, Johannes Bracher, Timo Dimitriadis, Veit Hagenmeyer, Alexander I. Jordan, Sebastian Lerch, Kaleb Phipps, and Melanie Schienle. Model Diagnostics and Forecast Evaluation for Quantiles. *Annual Review of Statistics and Its Application*, 10(1):597–621, March 2023. ISSN 2326-8298, 2326-831X. doi: 10.1146/annurev-statistics-032921-020240. URL https://www.annualreviews.org/doi/10.1146/annurev-statistics-032921-020240.

[649] Sam Allen, Christopher A. T. Ferro, and Frank Kwasniok. A conditional decomposition of proper scores: quantifying the sources of information in a forecast. *Quarterly Journal of the Royal Meteorological Society*, 149(754):1704–1725, July 2023. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.4478. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4478.

[650] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer, editors. *Algorithmic Learning in a Random World*. SpringerLink Bücher. Springer Science+Business Media, Inc, Boston, MA, 2005. ISBN 978-0-387-00152-4 978-0-387-25061-8. doi: 10.1007/b106715.

[651] Isaac Gibbs and Emmanuel Candès. Conformal Inference for Online Prediction with Arbitrary Distribution Shifts, October 2023. URL http://arxiv.org/abs/2208.08401. arXiv:2208.08401 [stat].

[652] Werner Ehm, Tilmann Gneiting, Alexander Jordan, and Fabian Krüger. Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations and Forecast Rankings. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(3):505–562, June 2016. ISSN 1369-7412, 1467-9868. doi: 10.1111/rssb.12154. URL https://academic.oup.com/jrsssb/article/78/3/505/7040984.

[653] Zied Ben Bouallègue, Thomas Haiden, and David S. Richardson. The diagonal score: Definition, properties, and interpretations. *Quarterly Journal of the Royal Meteorological Society*, 144(714):1463–1473, July 2018. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.3293. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3293.

[654] Tim Palmer and David Richardson. Decisions, decisions. . . ! 2014. doi: 10.21957/BYCHJ3CF. URL https://www.ecmwf.int/node/17333. Publisher: ECMWF.

[655] Erez Buchweitz, João Vitor Romano, and Ryan J. Tibshirani. Asymmetric Penalties Underlie Proper Loss Functions in Probabilistic Forecasting, May 2025. URL http://arxiv.org/abs/2505.00937. arXiv:2505.00937 [math].

[656] Thomas J. Kniesner and W. Kip Viscusi. The Value of a Statistical Life. In *Oxford Research Encyclopedia of Economics and Finance*. July 2019. ISBN 978-0-19-062597-9. doi: 10.1093/acrefore/9780190625979.013.138. URL https://oxfordre.com/economics/display/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-138.

[657] Christopher McCabe, Karl Claxton, and Anthony J. Culyer. The NICE cost-effectiveness threshold: what it is and what that means. *PharmacoEconomics*, 26(9):733–744, 2008. ISSN 1170-7690. doi: 10.2165/00019053-200826090-00004.

[658] Nicholas J. Irons and Adrian E. Raftery. US COVID-19 school closure was not cost-effective, but other measures were, November 2024. URL http://arxiv.org/abs/2411.12016. arXiv:2411.12016 [stat].

[659] Matt J. Keeling, Edward M. Hill, Stavros Petrou, Phuong Bich Tran, May Ee Png, Sophie Staniszewska, Corinna Clark, Katie Hassel, Julia Stowe, and Nick Andrews. Cost-effectiveness of routine COVID-19 adult vaccination programmes in England, November 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.11.08.24316972.

[660] Siyu Chen, Richard Creswell, Rachel Hounsell, Liberty Cantrell, Sumali Bajaj, Prabin Dahal, Joseph Tsui Lok Hei, Olumide Kolade, Ma'ayan Amswych, Reshania Naidoo, Tom Fowler, Susan Hopkins, Kasia Stepniewska, Merryn Voysey, Lisa White, Rima Shretta, and Ben Lambert. Mass testing for discovery and control of COVID-19 outbreaks in adult social care: an observational study and cost-effectiveness analysis of 14 805 care homes in England. *BMJ Public Health*, 3(1), March 2025. ISSN 2753-4294. doi: 10.1136/bmjph-2024-001376. URL https://bmjpublichealth.bmj.com/content/3/1/e001376. Publisher: BMJ Publishing Group Ltd.

[661] Hugo C. Turner, Duong Le Quyen, Reynold Dias, Phan Thi Huong, Cameron P. Simmons, and Katherine L. Anders. An economic evaluation of Wolbachia deployments for dengue control in Vietnam. *PLOS Neglected Tropical Diseases*, 17(5):e0011356, May 2023. ISSN 1935-2735. doi: 10.1371/journal.pntd.0011356. URL https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0011356. Publisher: Public Library of Science.

[662] Wu Zeng, Yara A. Halasa-Rappel, Nicolas Baurin, Laurent Coudeville, and Donald S. Shepard. Cost-effectiveness of dengue vaccination in ten endemic countries. *Vaccine*, 36(3):413–420, January 2018. ISSN 0264410X. doi: 10.1016/j.vaccine.2017.11.064. URL https://linkinghub.elsevier.com/retrieve/pii/S0264410X17316638.

[663] Timo Dimitriadis, Tilmann Gneiting, Alexander I. Jordan, and Peter Vogel. Evaluating probabilistic classifiers: The triptych. *International Journal of Forecasting*, 40(3):1101–1122, July 2024. ISSN 01692070. doi: 10.1016/j.ijforecast.2023.09.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0169207023000997.

[664] Tilmann Gneiting and Peter Vogel. Receiver operating characteristic (ROC) curves: equivalences, beta model, and minimum distance estimation. *Machine Learning*, 111(6):2147–2159, June 2022. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-021-06115-2. URL https://link.springer.com/10.1007/s10994-021-06115-2.

[665] R Tyrrell Rockafellar, Stanislav Uryasev, and others. Optimization of conditional value-at-risk. *Journal of risk*, 2: 21–42, 2000. Publisher: Citeseer.

[666] Hamed Rahimian and Sanjay Mehrotra. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3:1–85, 2022.

[667] Stefan Weber. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 16(2):419–441, 2006. Publisher: Wiley Online Library.

[668] Tobias Fissler and Johanna F. Ziegel. Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44 (4):1680 – 1707, 2016. doi: 10.1214/16-AOS1439. URL https://doi.org/10.1214/16-AOS1439. Publisher: Institute of Mathematical Statistics.

[669] Richard W. Katz and Jeffrey K. Lazo. Economic Value of Weather and Climate Forecasts. In Michael P. Clements and David F. Hendry, editors, *The Oxford Handbook of Economic Forecasting*, page 0. Oxford University Press, July 2011. ISBN 978-0-19-539864-9. doi: 10.1093/oxfordhb/9780195398649.013.0021. URL https://doi.org/10.1093/oxfordhb/9780195398649.013.0021.

[670] Alexis Robert, Lloyd A. C. Chapman, Rok Grah, Rene Niehus, Frank Sandmann, Bastian Prasse, Sebastian Funk, and Adam J. Kucharski. Predicting subnational incidence of COVID-19 cases and deaths in EU countries. *BMC Infectious Diseases*, 24(1):204, February 2024. ISSN 1471-2334. doi: 10.1186/s12879-024-08986-x. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-024-08986-x.

[671] Rob J. Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. Otexts, Online Open-Access Textbooks, Melbourne, Australia, third print edition edition, 2021. ISBN 978-0-9875071-3-6.

[672] Paul-Christian Bürkner, Jonah Gabry, and Aki Vehtari. Approximate leave-future-out cross-validation for Bayesian time series models. *Journal of Statistical Computation and Simulation*, 90(14):2499–2523, September 2020. ISSN 0094-9655, 1563-5163. doi: 10.1080/00949655.2020.1783262. URL https://www.tandfonline.com/doi/full/10.1080/00949655.2020.1783262.

[673] Tilmann Gneiting and Johannes Resin. Regression diagnostics meets forecast evaluation: conditional calibration, reliability diagrams, and coefficient of determination. *Electronic Journal of Statistics*, 17(2), January 2023. ISSN 1935-7524. doi: 10.1214/23-EJS2180. URL https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-17/issue-2/Regression-diagnostics-meets-forecast-evaluation--conditional-calibration-reliability-diagrams/10.1214/23-EJS2180.full.

[674] Cabinet Office. Independent review into the response to the 2009 swine flu pandemic. Technical report, Cabinet Office, July 2010.

[675] Charles Blackorby, Walter Bossert, and David J. Donaldson. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge University Press, 1 edition, August 2005. ISBN 978-0-521-53258-7 978-0-521-82551-1 978-1-139-05224-5. doi: 10.1017/CCOL0521825512. URL https://www.cambridge.org/core/product/identifier/9781139052245/type/book.

[676] Kenneth J. Arrow. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, 58(4):328–346, August 1950. ISSN 0022-3808, 1537-534X. doi: 10.1086/256963. URL https://www.journals.uchicago.edu/doi/10.1086/256963.

[677] Grace Liu, Jake C. Snell, Thomas L. Griffiths, and Rachit Dubey. Binary climate data visuals amplify perceived impact of climate change. *Nature Human Behaviour*, April 2025. ISSN 2397-3374. doi: 10.1038/s41562-025-02183-9. URL https://www.nature.com/articles/s41562-025-02183-9.

[678] Claudia R. Schneider, Alexandra L. J. Freeman, David Spiegelhalter, and Sander Van Der Linden. The effects of communicating scientific uncertainty on trust and decision making in a public health context. *Judgment and Decision Making*, 17(4):849–882, July 2022. ISSN 1930-2975. doi: 10.1017/S1930297500008962. URL https://www.cambridge.org/core/product/identifier/S1930297500008962/type/journal_article.

[679] Robert J. Taggart and David J. Wilke. Warnings based on risk matrices: a coherent framework with consistent evaluation, March 2025. URL https://egusphere.copernicus.org/preprints/2025/egusphere-2025-323/.

[680] Manit Sittimart, Waranya Rattanavipapong, Andrew J. Mirelman, Trinh Manh Hung, Saudamini Dabak, Laura E. Downey, Mark Jit, Yot Teerawattananon, and Hugo C. Turner. An overview of the perspectives used in health economic evaluations. *Cost Effectiveness and Resource Allocation*, 22(1):41, May 2024. ISSN 1478-7547. doi: 10.1186/s12962-024-00552-1. URL https://resource-allocation.biomedcentral.com/articles/10.1186/s12962-024-00552-1.

[681] M. J. Rodwell, J. Hammond, S. Thornton, and D. S. Richardson. User decisions, and how these could guide developments in probabilistic forecasting. *Quarterly Journal of the Royal Meteorological Society*, 146(732):3266–3284, October 2020. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.3845. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3845.

[682] Yuyang Chen, Naizhe Li, José Lourenço, Lin Wang, Bernard Cazelles, Lu Dong, Bingying Li, Yang Liu, Mark Jit, Nikos I Bosse, Sam Abbott, Raman Velayudhan, Annelies Wilder-Smith, Huaiyu Tian, Oliver J Brady, Simon R Procter, Kerry Lm Wong, Joel Hellewell, Nicholas G Davies, Christopher I Jarvis, Ciara V McCarthy, Graham Medley, Sophie R Meakin, Alicia Rosello, Emilie Finch, Rachel Lowe, Carl A B Pearson, Samuel Clifford, Billy J Quilty, Stefan Flasche, Hamish P Gibbs, Lloyd A C Chapman, Katherine E. Atkins, David Hodgson, Rosanna C Barnard, Timothy W Russell, Petra Klepac, Yalda Jafari, Rosalind M Eggo, Paul Mee, Matthew Quaife, Akira Endo, Sebastian Funk, Stéphane Hué, Adam J Kucharski, W John Edmunds, Kathleen O'Reilly, Rachael Pung, C Julian Villabona-Arenas, Amy Gimma, Kaja Abbas, Kiesha Prem, Gwenan M Knight, Fiona Yueqian Sun, William Waites, James D Munday, Mihaly Koltai, Frank G Sandmann, and Damien C Tully. Measuring the effects of COVID-19-related disruption on dengue transmission in southeast Asia and Latin America: a statistical modelling study. *The Lancet Infectious Diseases*, 22(5):657–667, May 2022. ISSN 14733099. doi: 10.1016/S1473-3099(22)00025-1. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309922000251.

[683] Alexander E. Zarebski, Nefel Tellioglu, Jessica E. Stockdale, Julie A. Spencer, Wasiur R. KhudaBukhsh, Joel C. Miller, and Cameron Zachreson. Including frameworks of public health ethics in computational modelling of infectious disease interventions, January 2025. URL http://arxiv.org/abs/2502.00071. arXiv:2502.00071 [physics].

[684] Velma K. Lopez, Leonardo S. Bastos, Cláudia T. Codeço, and Michael A. Johansson. A roadmap to account for reporting delays for public health situational awareness – a case study with COVID-19 and dengue in United States jurisdictions, November 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.11.09.24315999.

[685] Yo Joong Choe and Aaditya Ramdas. Comparing Sequential Forecasters. *Operations Research*, 72(4):1368–1387, July 2024. ISSN 0030-364X. doi: 10.1287/opre.2021.0792. URL https://pubsonline.informs.org/doi/10.1287/opre.2021.0792. Publisher: INFORMS.

[686] Axel Gandy, Kaushik Jana, and Almut E. D. Veraart. Scoring predictions at extreme quantiles. *AStA Advances in Statistical Analysis*, 106(4):527–544, December 2022. ISSN 1863-818X. doi: 10.1007/s10182-021-00421-9. URL https://doi.org/10.1007/s10182-021-00421-9.

[687] Raffaella Giacomini and Ivana Komunjer. Evaluation and Combination of Conditional Quantile Forecasts. *Journal of Business & Economic Statistics*, 23(4):416–431, October 2005. ISSN 0735-0015. doi: 10.1198/073500105000000018. URL https://doi.org/10.1198/073500105000000018. Publisher: ASA Website _eprint: https://doi.org/10.1198/073500105000000018.

[688] John Stewart Fabila-Carrasco, Chao Tan, and Javier Escudero. Multivariate permutation entropy, a Cartesian graph product approach, March 2022. URL http://arxiv.org/abs/2203.00550. arXiv:2203.00550 [math].

[689] Xin Huang, Han Lin Shang, and David Pitt. Permutation entropy and its variants for measuring temporal dependence. *Australian & New Zealand Journal of Statistics*, 64(4):442–477, December 2022. ISSN 1369-1473, 1467-842X. doi: 10.1111/anzs.12376. URL https://onlinelibrary.wiley.com/doi/10.1111/anzs.12376.

[690] Ian Lundberg, Rachel Brown-Weinstock, Susan Clampet-Lundquist, Sarah Pachman, Timothy J. Nelson, Vicki Yang, Kathryn Edin, and Matthew J. Salganik. The origins of unpredictability in life outcome prediction tasks. *Proceedings of the National Academy of Sciences*, 121(24):e2322973121, June 2024. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2322973121. URL https://pnas.org/doi/10.1073/pnas.2322973121.

[691] Joacim Rocklöv and Robert Dubrow. Climate change: an enduring challenge for vector-borne disease prevention and control. *Nature Immunology*, 21(5):479–483, May 2020. ISSN 1529-2908, 1529-2916. doi: 10.1038/s41590-020-0648-y. URL https://www.nature.com/articles/s41590-020-0648-y.

[692] Joseph L.-H. Tsui, Rosario Evans Pena, Monika Moir, Rhys P. D. Inward, Eduan Wilkinson, James Emmanuel San, Jenicca Poongavanan, Sumali Bajaj, Bernardo Gutierrez, Abhishek Dasgupta, Tulio De Oliveira, Moritz U. G. Kraemer, Houriiyah Tegally, and Prathyush Sambaturu. Impacts of climate change-related human migration on infectious diseases. *Nature Climate Change*, 14(8):793–802, August 2024. ISSN 1758-678X, 1758-6798. doi: 10.1038/s41558-024-02078-z. URL https://www.nature.com/articles/s41558-024-02078-z.

[693] Paolo Cornale, Michele Tizzani, Fabio Ciulla, Kyriaki Kalimeri, Elisa Omodei, Daniela Paolotti, and Yelena Mejova. The role of science in the climate change discussions on Reddit. *PLOS Climate*, 4(5):e0000541, May 2025. ISSN 2767-3200. doi: 10.1371/journal.pclm.0000541. URL https://journals.plos.org/climate/article?id=10.1371/journal.pclm.0000541. Publisher: Public Library of Science.

[694] Julian Kauk, Helene Kreysa, and Stefan R. Schweinberger. Understanding and countering the spread of conspiracy theories in social networks: Evidence from epidemiological models of Twitter data. *PLoS ONE*, 16(8):e0256179, August 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0256179. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8360523/.

[695] Richard A. Stein, Oana Ometa, Sarah Pachtman Shetty, Adi Katz, Mircea Ionut Popitiu, and Robert Brotherton. Conspiracy theories in the era of COVID-19: A tale of two pandemics. *International Journal of Clinical Practice*, 75 (2):e13778, February 2021. ISSN 1368-5031. doi: 10.1111/ijcp.13778. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7995222/.

[696] Jeanette L. Drake, Yekaterina Y. Kontar, John C. Eichelberger, T. Scott Rupp, and Karen M. Taylor, editors. *Communicating Climate-Change and Natural Hazard Risk and Cultivating Resilience: Case Studies for a Multi-disciplinary Approach*. Number 45 in Advances in Natural and Technological Hazards Research. Springer, Cham, 2016. ISBN 978-3-319-20160-3 978-3-319-20161-0. doi: 10.1007/978-3-319-20161-0.

[697] Gila Neta, William Pan, Kristie Ebi, Daniel F Buss, Trisha Castranio, Rachel Lowe, Sadie J Ryan, Anna M Stewart-Ibarra, Limb K Hapairai, Meena Sehgal, Michael C Wimberly, Leslie Rollock, Maureen Lichtveld, and John Balbus. Advancing climate change health adaptation through implementation science. *The Lancet Planetary Health*, 6(11):e909–e918, November 2022. ISSN 25425196. doi: 10.1016/S2542-5196(22)00199-1. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519622001991.

[698] World Health Organization. Report on the global arbovirus surveillance and response capacity survey 2021-2022. Technical report, March 2025. URL https://www.who.int/publications/i/item/9789240107380.

[699] Ayoade Alakija. Leveraging lessons from the COVID-19 pandemic to strengthen low-income and middle-income country preparedness for future global health threats. *The Lancet Infectious Diseases*, 23(8):e310–e317, August 2023. ISSN 14733099. doi: 10.1016/S1473-3099(23)00279-7. URL https://linkinghub.elsevier.com/retrieve/pii/S1473309923002797.

[700] Christopher Jackson, Anne Presanis, Stefano Conti, and Daniela De Angelis. Value of Information: Sensitivity Analysis and Research Design in Bayesian Evidence Synthesis. *Journal of the American Statistical Association*, 114(528):1436–1449, October 2019. ISSN 0162-1459, 1537-274X. doi: http://refhub.elsevier.com/S0169-2070(23)00099-7/sb65. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1562932.

[701] Wei-Xian Zhang, Tian-Yu Zhao, Cun-Chen Wang, Yong He, Hong-Zheng Lu, Hai-Ting Zhang, Lin-Min Wang, Mao Zhang, Chun-Xiao Li, and Sheng-Qun Deng. Assessing the global dengue burden: Incidence, mortality, and disability trends over three decades. *PLOS Neglected Tropical Diseases*, 19(3):e0012932, March 2025. ISSN 1935-2735. doi: 10.1371/journal.pntd.0012932. URL https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0012932. Publisher: Public Library of Science.

[702] Leigh R. Bowman, Silvia Runge-Ranzinger, and P. J. McCall. Assessing the relationship between vector indices and dengue transmission: a systematic review of the evidence. *PLoS neglected tropical diseases*, 8(5):e2848, May 2014. ISSN 1935-2735. doi: 10.1371/journal.pntd.0002848.

[703] Lee Ching Ng. Challenges in dengue surveillance and control. *Western Pacific Surveillance and Response Journal : WPSAR*, 2(2):1–3, April 2011. ISSN 2094-7321. doi: 10.5365/WPSAR.2011.2.2.001. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3730958/.

[704] Emily Mary Gainor, Eva Harris, and A. Desiree LaBeaud. Uncovering the Burden of Dengue in Africa: Considerations on Magnitude, Misdiagnosis, and Ancestry. *Viruses*, 14(2):233, February 2022. ISSN 1999-4915. doi: 10.3390/v14020233. URL https://www.mdpi.com/1999-4915/14/2/233. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[705] Jan M. Brauner, Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B. Stephenson, Gavin Leech, George Altman, Vladimir Mikulik, Alexander John Norman, Joshua Teperowski Monrad, Tamay Besiroglu, Hong Ge, Meghan A. Hartwick, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal, and Jan Kulveit. Inferring the effectiveness of government interventions against COVID-19. *Science*, 371(6531):eabd9338, February 2021. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abd9338. URL https://www.science.org/doi/10.1126/science.abd9338.

[706] Alfred J. Lotka. Quantitative Studies in Epidemiology. *Nature*, 88(2206):497–498, February 1912. ISSN 0028-0836, 1476-4687. doi: 10.1038/088497b0. URL https://www.nature.com/articles/088497b0.

[707] Michael B. Bonsall. Vector Control, Optimal Control, and Vector-Borne Disease Dynamics. In John M. Drake, Michael Bonsall, and Michael Strand, editors, *Population Biology of Vector-Borne Diseases*, pages 267–288. Oxford University PressOxford, 1 edition, December 2020. ISBN 978-0-19-885324-4 978-0-19-188771-0. doi: 10.1093/oso/9780198853244.003.0015. URL https://academic.oup.com/book/41101/chapter/350400469.

[708] Thomas Rawson, Kym E. Wilkins, and Michael B. Bonsall. Optimal control approaches for combining medicines and mosquito control in tackling dengue. *Royal Society Open Science*, 7(4):181843, April 2020. ISSN 2054-5703. doi: 10.1098/rsos.181843. URL https://royalsocietypublishing.org/doi/10.1098/rsos.181843.

[709] Ellie Sherrard-Smith, Peter Winskill, Arran Hamlet, Corine Ngufor, Raphael N'Guessan, Moussa W Guelbeogo, Antoine Sanou, Rebecca K Nash, Alexander Hill, Emma L Russell, Mark Woodbridge, Patrick K Tungu, Mara D Kont, Tom Mclean, Christen Fornadel, Jason H Richardson, Martin J Donnelly, Sarah G Staedke, Samuel Gonahasa, Natacha Protopopoff, Mark Rowland, and Thomas S Churcher. Optimising the deployment of vector control tools against malaria: a data-informed modelling study. *The Lancet Planetary Health*, 6(2):e100–e109, February 2022. ISSN 2542-5196. doi: 10.1016/S2542-5196(21)00296-5. URL https://doi.org/10.1016/S2542-5196(21)00296-5. Publisher: Elsevier.

[710] Meksianis Z. Ndii, R.I. Hickson, David Allingham, and G.N. Mercer. Modelling the transmission dynamics of dengue in the presence of Wolbachia. *Mathematical Biosciences*, 262:157–166, April 2015. ISSN 00255564. doi: 10.1016/j.mbs.2014.12.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0025556415000164.

[711] Pan American Health Organization. *Integrated Management Strategy for Arboviral Disease Prevention and Control in the Americas*. Organización Panamericana de la Salud, 2020. ISBN 978-92-75-12049-1. doi: 10.37774/9789275120491. URL https://iris.paho.org/handle/10665.2/52492.

[712] Sebastian Funk, Iza Ciglenecki, Amanda Tiffany, Etienne Gignoux, Anton Camacho, Rosalind M. Eggo, Adam J. Kucharski, W. John Edmunds, Josephus Bolongei, Phillip Azuma, Peter Clement, Tamba S. Alpha, Esther Sterk, Barbara Telfer, Gregory Engel, Lucy Anne Parker, Motoi Suzuki, Nico Heijenberg, and Bruce Reeder. The impact of control strategies and behavioural changes on the elimination of Ebola from Lofa County, Liberia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1721):20160302, May 2017. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2016.0302. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0302.

[713] Giovanni Spitale. COVID-19 and the ethics of quarantine: a lesson from the Eyam plague. *Medicine, Health Care and Philosophy*, 23(4):603–609, December 2020. ISSN 1386-7423, 1572-8633. doi: 10.1007/s11019-020-09971-2. URL https://link.springer.com/10.1007/s11019-020-09971-2.

[714] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, December 1999. ISSN 0737-4038, 1537-1719. doi: 10.1093/oxfordjournals.molbev.a026091. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a026091.

[715] S. A. Sisson, Y. Fan, and Mark M. Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, February 2007. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0607208104. URL https://pnas.org/doi/full/10.1073/pnas.0607208104.

[716] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P.H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, February 2009. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2008.0172. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2008.0172.

[717] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, February 2019. ISSN 00219991. doi: 10.1016/j.jcp.2018.10.045. URL https://linkinghub.elsevier.com/retrieve/pii/S0021999118307125.

[718] Pierre-alexandre Kamienny, Stéphane d' Ascoli, Guillaume Lample, and Francois Charton. End-to-end Symbolic Regression with Transformers. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 10269–10281. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/42eb37cdbefd7abae0835f4b67548c39-Paper-Conference.pdf.

[719] C. Garrett-Jones and G. R. Shidrawi. Malaria vectorial capacity of a population of Anopheles gambiae: an exercise in epidemiological entomology. *Bulletin of the World Health Organization*, 40(4):531–545, April 1969. ISSN 0042-9686.

[720] Daniele Da Re, Veronica Andreo, Tomas San Miguel, Margo Blaha, Roberto Rosà, Annapaola Rizzoli, Joe Harrison, Sean Sorek, Leah R. Johnson, and Paul J. Huxley. AedesTraits: A global database of temperature–dependent trait responses in Aedes mosquitoes. April 2025. URL https://ecoevorxiv.org/repository/view/8919/. Publisher: EcoEvoRxiv.

[721] Edward M. Hill, Ryan , Matthew, Haw , David, Lynch , Mark P., McCabe , Ruth, Milne , Alice E., Turner , Matthew S., Vedhara , Kavita, Zeng , Fanqi, Barons , Martine J., Nixon , Emily J., Parnell , Stephen, , and Kirsty J. Bolton. Integrating human behaviour and epidemiological modelling: unlocking the remaining challenges. *Mathematics in Medical and Life Sciences*, 1(1):2429479, December 2024. ISSN null. doi: 10.1080/29937574.2024.2429479. URL https://doi.org/10.1080/29937574.2024.2429479. Publisher: Taylor & Francis Open _eprint: https://doi.org/10.1080/29937574.2024.2429479.

[722] Carl T. Bergstrom and William P. Hanage. Human behavior and disease dynamics. *Proceedings of the National Academy of Sciences*, 121(1):e2317211120, January 2024. doi: 10.1073/pnas.2317211120. URL https://www.pnas.org/doi/10.1073/pnas.2317211120. Publisher: Proceedings of the National Academy of Sciences.

[723] Tijani A. Sulaimon, Anthony J. Wood, Michael B. Bonsall, Michael Boots, and Jennifer S. Lord. The potential impacts of vector host species fidelity on zoonotic arbovirus transmission. *PLoS neglected tropical diseases*, 19(5): e0012196, May 2025. ISSN 1935-2735. doi: 10.1371/journal.pntd.0012196.

[724] René Saameli, Dikolela Kalubi, Mark Herringer, Tim Sutton, and Eric de Roodenbeke. Healthsites.io: The Global Healthsites Mapping Project. In Silvia Hostettler, Samira Najih Besson, and Jean-Claude Bolay, editors, *Technologies for Development*, pages 53–59, Cham, 2018. Springer International Publishing. ISBN 978-3-319-91068-0. doi: 10.1007/978-3-319-91068-0_5.

[725] José Lourenço, Warren Tennant, Nuno R. Faria, Andrew Walker, Sunetra Gupta, and Mario Recker. Challenges in dengue research: A computational perspective. *Evolutionary Applications*, 11(4):516–533, 2018. ISSN 1752-4571. doi: 10.1111/eva.12554. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.12554. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/eva.12554.

[726] Kris V. Parag and Patrick Hoscheit. Why do border controls fail to stabilise epidemics?, May 2025. URL https://www.medrxiv.org/content/10.1101/2025.05.12.25327440v1. Pages: 2025.05.12.25327440.

[727] Mousumi Roy, Hannah E. Clapham, and Swapnil Mishra. Incorporating human mobility to enhance epidemic response and estimate real-time reproduction numbers, April 2025. URL http://medrxiv.org/lookup/doi/10.1101/2025.04.25.25326411.

[728] Yuli Zhang, Xiao Liu, Zhen Wu, Shuo Feng, Ke Lu, Wenbing Zhu, Hengyi Sun, and Guoyu Niu. Oropouche virus: A neglected global arboviral threat. *Virus Research*, 341:199318, March 2024. ISSN 01681702. doi: 10.1016/j.virusres.2024.199318. URL https://linkinghub.elsevier.com/retrieve/pii/S016817022400011X.

[729] Arcady Ponosov, Lev Idels, and Ramazan Kadiev. Stochastic McKendrick–Von Foerster models with applications. *Physica A: Statistical Mechanics and its Applications*, 537:122641, January 2020. ISSN 03784371. doi: 10.1016/j.physa.2019.122641. URL https://linkinghub.elsevier.com/retrieve/pii/S0378437119315092.

[730] Maxim A. Alshanskiy. A model of age-structured population under stochastic perturbation of death and birth rates. *Ural mathematical journal*, 4(1):3–13, 2018. ISSN 24143952. doi: 10.15826/umj.2018.1.001. URL https://umjuran.ru/index.php/umj/article/view/106.

[731] Zhang Qimin and Li Xining. Existence and Uniqueness for Stochastic Age-Dependent Population with Fractional Brownian Motion. *Mathematical Problems in Engineering*, 2012(1):813535, January 2012. ISSN 1024-123X, 1563-5147. doi: 10.1155/2012/813535. URL https://onlinelibrary.wiley.com/doi/10.1155/2012/813535.

[732] Kenny S Crump and Charles J Mode. A general age-dependent branching process. I. *Journal of Mathematical Analysis and Applications*, 24(3):494–508, December 1968. ISSN 0022247X. doi: 10.1016/0022-247X(68)90005-X. URL https://linkinghub.elsevier.com/retrieve/pii/0022247X6890005X.

[733] Kenny Crump and Charles J Mode. A general age-dependent branching process. II. *Journal of Mathematical Analysis and Applications*, 25(1):8–17, January 1969. ISSN 0022247X. doi: 10.1016/0022-247X(69)90210-8. URL https://linkinghub.elsevier.com/retrieve/pii/0022247X69902108.

[734] Peter Jagers. *Branching processes with biological applications*. Wiley series in probability and mathematical statistics. Wiley, London, 1975. ISBN 978-0-471-43652-2.

[735] E. Southall, Z. Ogi-Gittins, A.R. Kaye, W.S. Hart, F.A. Lovell-Read, and R.N. Thompson. A practical guide to mathematical methods for estimating infectious disease outbreak risks. *Journal of Theoretical Biology*, 562:111417, April 2023. ISSN 00225193. doi: 10.1016/j.jtbi.2023.111417. URL https://linkinghub.elsevier.com/retrieve/pii/S0022519323000139.

[736] Sunil Maity and Partha Sarathi Mandal. A Comparison of Deterministic and Stochastic Plant-Vector-Virus Models Based on Probability of Disease Extinction and Outbreak. *Bulletin of Mathematical Biology*, 84(3):41, March 2022. ISSN 0092-8240, 1522-9602. doi: 10.1007/s11538-022-01001-x. URL https://link.springer.com/10.1007/s11538-022-01001-x.

[737] Asma Akter Akhi, Kazi Mehedi Mohammad, and Md. Kamrujjaman. Seasonal variability and stochastic branching process in malaria outbreak probability. *Acta Tropica*, 257:107311, September 2024. ISSN 0001706X. doi: 10.1016/j.actatropica.2024.107311. URL https://linkinghub.elsevier.com/retrieve/pii/S0001706X24001931.

[738] Alun L Lloyd, Ji Zhang, and A.Morgan Root. Stochasticity and heterogeneity in host–vector models. *Journal of The Royal Society Interface*, 4(16):851–863, October 2007. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2007.1064. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2007.1064.

[739] Adam J. Kucharski. Controlling minor outbreaks is necessary to prepare for major pandemics. *PLOS Biology*, 22 (12):e3002945, December 2024. ISSN 1545-7885. doi: 10.1371/journal.pbio.3002945. URL https://dx.plos.org/10.1371/journal.pbio.3002945.

[740] Sandra Barman, Jan C. Semenza, Pratik Singh, Henrik Sjödin, Joacim Rocklöv, and Jonas Wallin. A climate and population dependent diffusion model forecasts the spread of Aedes Albopictus mosquitoes in Europe. *Communications Earth & Environment*, 6(1):276, April 2025. ISSN 2662-4435. doi: 10.1038/s43247-025-02199-z. URL https://www.nature.com/articles/s43247-025-02199-z.

[741] Davey L. Jones, Jennifer M. Rhymes, Matthew J. Wade, Jessica L. Kevill, Shelagh K. Malham, Jasmine M.S. Grimsley, Charlotte Rimmer, Andrew J. Weightman, and Kata Farkas. Suitability of aircraft wastewater for pathogen detection and public health surveillance. *Science of The Total Environment*, 856:159162, January 2023. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2022.159162. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969722062611. Publisher: Elsevier BV.

[742] Guillaume St-Onge, Jessica T. Davis, Laurent Hébert-Dufresne, Antoine Allard, Alessandra Urbinati, Samuel V. Scarpino, Matteo Chinazzi, and Alessandro Vespignani. Pandemic monitoring with global aircraft-based wastewater surveillance networks. *Nature Medicine*, 31(3):788–796, March 2025. ISSN 1546-170X. doi: 10.1038/s41591-025-03501-4. URL https://www.nature.com/articles/s41591-025-03501-4. Publisher: Nature Publishing Group.

[743] Andrew Bo Liu, Daniel Lee, Amogh Prabhav Jalihal, William P. Hanage, and Michael Springer. Quantitatively assessing early detection strategies for mitigating COVID-19 and future pandemics. *Nature Communications*, 14 (1):8479, December 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-44199-7. URL https://www.nature.com/articles/s41467-023-44199-7. Publisher: Nature Publishing Group.

[744] Carmen Tamayo Cuartero, Anna C. Carnegie, Zulma M. Cucunuba, Anne Cori, Sara M. Hollis, Rolina D. Van Gaalen, Amrish Y. Baidjoe, Alexander F. Spina, John A. Lees, Simon Cauchemez, Mauricio Santos, Juan D. Umaña, Chaoran Chen, Hugo Gruson, Pratik Gupte, Joseph Tsui, Anita A. Shah, Geraldine Gomez Millan, David Santiago Quevedo, Neale Batra, Andrea Torneri, and Adam J. Kucharski. From the 100 Day Mission to 100 lines of software

development: how to improve early outbreak analytics. *The Lancet Digital Health*, 7(2):e161–e166, February 2025. ISSN 2589-7500. doi: 10.1016/S2589-7500(24)00218-8. URL https://www.thelancet.com/journals/landig/article/PIIS2589-7500(24)00218-8/fulltext. Publisher: Elsevier.

[745] K.A. Hamilton, M.J. Wade, K.G. Barnes, R.A. Street, and S. Paterson. Wastewater-based epidemiology as a public health resource in low- and middle-income settings. *Environmental Pollution*, 351:124045, June 2024. ISSN 02697491. doi: 10.1016/j.envpol.2024.124045. URL https://linkinghub.elsevier.com/retrieve/pii/S0269749124007590.

[746] Aparna Keshaviah, Megan B Diamond, Matthew J Wade, Samuel V Scarpino, Warish Ahmed, Fabian Amman, Olusola Aruna, Andrei Badilla-Aguilar, Itay Bar-Or, Andreas Bergthaler, Julie E Bines, Aaron W Bivins, Alexandria B Boehm, Jean-Martin Brault, Jean-Baptiste Burnet, Joanne R Chapman, Angela Chaudhuri, Ana Maria De Roda Husman, Robert Delatolla, John J Dennehy, Megan Beth Diamond, Celeste Donato, Erwin Duizer, Abiodun Egwuenu, Oran Erster, Despo Fatta-Kassinos, Aldo Gaggero, Deirdre F Gilpin, Brent J Gilpin, Tyson E Graber, Christopher A Green, Amanda Handley, Joanne Hewitt, Rochelle H Holm, Heribert Insam, Marc C Johnson, Rabia Johnson, Davey L Jones, Timothy R Julian, Asha Jyothi, Aparna Keshaviah, Tamar Kohn, Katrin G Kuhn, Giuseppina La Rosa, Marie Lesenfants, Douglas G Manuel, Patrick M D'Aoust, Rudolf Markt, John W McGrath, Gertjan Medema, Christine L Moe, Indah Kartika Murni, Humood Naser, Colleen C Naughton, Leslie Ogorzaly, Vicka Oktaria, Christoph Ort, Popi Karaolia, Ekta H Patel, Steve Paterson, Mahbubur Rahman, Pablo Rivera-Navarro, Alex Robinson, Monica C Santa-Maria, Samuel V Scarpino, Heike Schmitt, Theodore Smith, Lauren B Stadler, Jorgen Stassijns, Alberta Stenico, Renee A Street, Elisabetta Suffredini, Zachary Susswein, Monica Trujillo, Matthew J Wade, Marlene K Wolfe, Habib Yakubu, and Maria Ines Zanoli Sato. Wastewater monitoring can anchor global disease surveillance systems. *The Lancet Global Health*, 11(6):e976–e981, June 2023. ISSN 2214109X. doi: 10.1016/S2214-109X(23)00170-5. URL https://linkinghub.elsevier.com/retrieve/pii/S2214109X23001705.

[747] Sindhu Ravuri, Elisabeth Burnor, Isobel Routledge, Natalie M. Linton, Mugdha Thakur, Alexandria Boehm, Marlene Wolfe, Heather N. Bischel, Colleen C. Naughton, Alexander T. Yu, Lauren A. White, and Tomás M. León. Estimating effective reproduction numbers using wastewater data from multiple sewersheds for SARS-CoV-2 in California counties. *Epidemics*, 50:100803, March 2025. ISSN 1755-4365. doi: 10.1016/j.epidem.2024.100803. URL https://www.sciencedirect.com/science/article/pii/S1755436524000641.

[748] Hiroki Ando and Kelly A. Reynolds. Wastewater-based effective reproduction number and prediction under the absence of shedding information. *Environment International*, 194:109128, December 2024. ISSN 01604120. doi: 10.1016/j.envint.2024.109128. URL https://linkinghub.elsevier.com/retrieve/pii/S0160412024007141.

[749] Jana S. Huisman, Jérémie Scire, Lea Caduff, Xavier Fernandez-Cassi, Pravin Ganesanandamoorthy, Anina Kull, Andreas Scheidegger, Elyse Stachler, Alexandria B. Boehm, Bridgette Hughes, Alisha Knudson, Aaron Topol, Krista R. Wigginton, Marlene K. Wolfe, Tamar Kohn, Christoph Ort, Tanja Stadler, and Timothy R. Julian. Wastewater-Based Estimation of the Effective Reproductive Number of SARS-CoV-2. *Environmental Health Perspectives*, 130(5):57011, May 2022. ISSN 1552-9924. doi: 10.1289/EHP10050.

[750] Leighton M. Watson, Michael J. Plank, Bridget A. Armstrong, Joanne R. Chapman, Joanne Hewitt, Helen Morris, Alvaro Orsi, Michael Bunce, Christl A. Donnelly, and Nicholas Steyn. Jointly estimating epidemiological dynamics of Covid-19 from case and wastewater data in Aotearoa New Zealand. *Communications Medicine*, 4(1): 1–9, July 2024. ISSN 2730-664X. doi: 10.1038/s43856-024-00570-3. URL https://www.nature.com/articles/s43856-024-00570-3. Publisher: Nature Publishing Group.

[751] Adrian Lison. adrian-lison/EpiSewer: EpiSewer v0.0.3, October 2024. URL https://zenodo.org/doi/10.5281/zenodo.13899759.

[752] Noman Bashir, Priya Donti, James Cuff, Sydney Sroka, Marija Ilic, Vivienne Sze, Christina Delimitrou, and Elsa Olivetti. The Climate and Sustainability Implications of Generative AI. *An MIT Exploration of Generative AI*, March 2024. URL https://mit-genai.pubpub.org/pub/8ulgrckc/release/2. Publisher: MIT.

[753] Paul A. Estabrooks, Ross C. Brownson, and Nicolaas P. Pronk. Dissemination and Implementation Science for Public Health Professionals: An Overview and Call to Action. *Preventing Chronic Disease*, 15:180525, December 2018. ISSN 1545-1151. doi: 10.5888/pcd15.180525. URL http://www.cdc.gov/pcd/issues/2018/18_0525.htm.

[754] Gilberto Sánchez-González and Renaud Condé. Mathematical modeling of Dengue virus serotypes propagation in Mexico. *PLOS ONE*, 18(7):1–13, July 2023. doi: 10.1371/journal.pone.0288392. URL https://doi.org/10.1371/journal.pone.0288392. Publisher: Public Library of Science.

[755] David Spiegelhalter. Why probability probably doesn't exist (but it is useful to act like it does). *Nature*, 636 (8043):560–563, December 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/d41586-024-04096-5. URL https://www.nature.com/articles/d41586-024-04096-5.

[756] Mandeep K Dhami and David R. Mandel. UK and US policies for communicating probability in intelligence analysis: A review, May 2020. URL https://osf.io/kuyhb.

[757] Galit Shmueli. To Explain or to Predict? *Statistical Science*, 25(3), August 2010. ISSN 0883-4237. doi: 10.1214/10-STS330. URL https://projecteuclid.org/journals/statistical-science/volume-25/issue-3/To-Explain-or-to-Predict/10.1214/10-STS330.full.

[758] Kari S McLeod. Our sense of Snow: the myth of John Snow in medical geography. *Social Science & Medicine*, 50(7-8):923–935, April 2000. ISSN 02779536. doi: 10.1016/S0277-9536(99)00345-7. URL https://linkinghub.elsevier.com/retrieve/pii/S0277953699003457.

[759] Thomas S. Coleman. Re-evaluating John Snow's 1856 south London study. *Social Science & Medicine*, 344:116612, March 2024. ISSN 02779536. doi: 10.1016/j.socscimed.2024.116612. URL https://linkinghub.elsevier.com/retrieve/pii/S027795362400056X.

[760] Reid C. Robson, Sonia M. Thomas, Étienne V. Langlois, Rhona Mijumbi, Ismael Kawooya, Jesmin Antony, Melissa Courvoisier, Krystle Amog, Robert Marten, Ivdity Chikovani, Devaki Nambiar, Rajani R. Ved, Soumyadeep Bhaumik, Nur Zahirah Balqis-Ali, Sondi Sararaks, Shakirah Md. Sharif, Rugare Abigail Kangwende, Ronald Munatsi, Sharon E. Straus, and Andrea C. Tricco. Embedding rapid reviews in health policy and systems decision-making: Impacts and lessons learned from four low- and middle-income countries. *Health Research Policy and Systems*, 21(1):45, June 2023. ISSN 1478-4505. doi: 10.1186/s12961-023-00992-w. URL https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-023-00992-w.

[761] Collins J Owek, Fatuma Hassan Guleid, Justinah Maluni, Joyline Jepkosgei, Vincent O Were, So Yoon Sim, Raymond Cw Hutubessy, Brittany L Hagedorn, Jacinta Nzinga, and Jacquie Oliwa. Lessons learned from COVID-19 modelling efforts for policy decision-making in lower- and middle-income countries. *BMJ Global Health*, 9(11): e015247, November 2024. ISSN 2059-7908. doi: 10.1136/bmjgh-2024-015247. URL https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2024-015247.

[762] Paul Cairney and Richard Kwiatkowski. How to communicate effectively with policymakers: combine insights from psychology and policy studies. *Palgrave Communications*, 3(1):37, November 2017. ISSN 2055-1045. doi: 10.1057/s41599-017-0046-8. URL https://www.nature.com/articles/s41599-017-0046-8.

[763] {Stan Development Team}. Stan Modeling Language Users Guide 2023 and Reference Manual, 2.3.1. https://mc-stan.org. URL https://mc-stan.org.

[764] Evan L. Ray and Nicholas G. Reich. Prediction of infectious disease epidemics via weighted density ensembles. *PLOS Computational Biology*, 14(2):e1005910, February 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005910. URL https://dx.plos.org/10.1371/journal.pcbi.1005910.

[765] Hiroshi Nishiura and Scott B. Halstead. Natural history of dengue virus (DENV)–1 and DENV-4 infections: reanalysis of classic studies. *The Journal of Infectious Diseases*, 195(7):1007–1013, April 2007. ISSN 0022-1899, 1537-6613. doi: 10.1086/511825. URL https://academic.oup.com/jid/article-lookup/doi/10.1086/511825.

[766] Alexander Jordan, Fabian Krüger, and Sebastian Lerch. Evaluating Probabilistic Forecasts with **scoringRules**. *Journal of Statistical Software*, 90(12), 2019. ISSN 1548-7660. doi: 10.18637/jss.v090.i12. URL http://www.jstatsoft.org/v90/i12/.

[767] Timo Dimitriadis and Alexander I. Jordan. triptych: Diagnostic Graphics to Evaluate Forecast Performance, October 2023. URL https://CRAN.R-project.org/package=triptych. Institution: Comprehensive R Archive Network Pages: 0.1.3.

Appendix A

Appendix — The utility of wastewater surveillance for monitoring SARS-CoV-2 prevalence

A.1 Relationship Between SARS-CoV-2 RNA Wastewater Concentrations and Prevalence

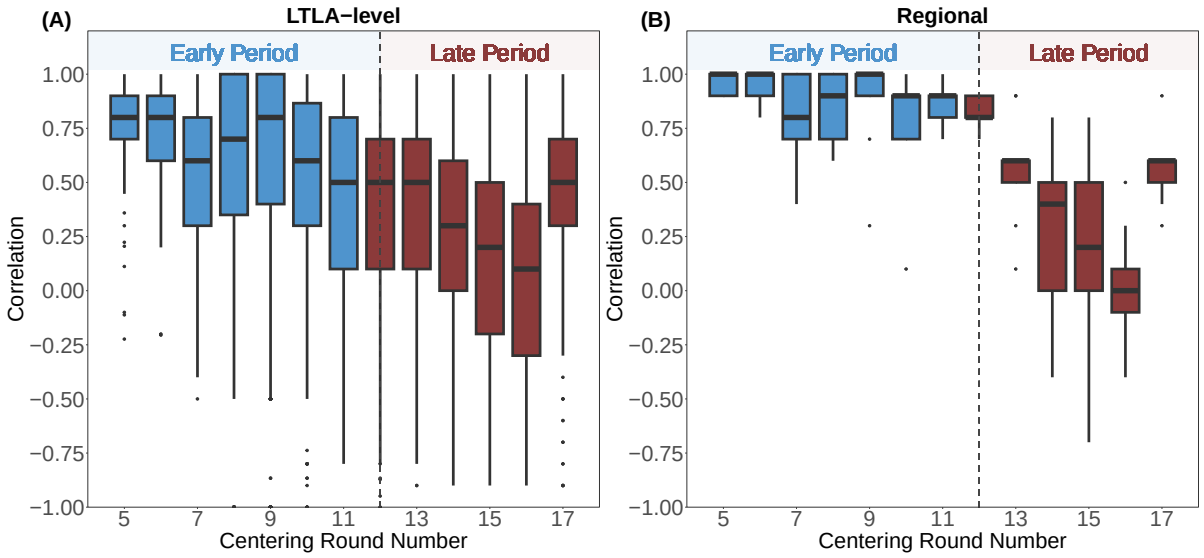


Figure SI A.1: **Prevalence-wastewater relationship within LTLAs and Regions.** This figure is similar to Figure 1 B-C, but here we compute Spearman's correlation between SARS-CoV-2 wastewater concentrations and REACT-1 SARS-CoV-2 prevalence for each LTLA and each region separately. Again, correlations were calculated within moving five-round windows centred on each survey round from REACT-1 rounds 5 (18 September 2020) to 17 (5 January 2022). Then, we present the results in boxplot visualisations, where the upper and lower hinges represent the first and third quartiles and the horizontal solid black line represents the median correlation. Whilst there are $n = 9$ regional correlations for each 5-round window, the number of LTLA-level correlations varies by survey round depending on the number of LTLAs included in the wastewater surveillance programme.

Table SI A.1: **Correlation analysis, by period, for wastewater concentrations and prevalence.** Spearman's correlation (r) is measured across all observations for each study period, where n is the number of observations used to calculate the corresponding correlation at each spatial resolution and a 95% CI is approximated using a paired bootstrap of 1000 replicates. Among LTLAs which report at least three wastewater concentration measurements across the early period - rounds 3 to 11 (24 July 2020 to 3 May 2021), the median LTLA-level correlation is 0.72, calculated as the median of each LTLA's individual correlation between wastewater concentrations and prevalence levels across the rounds. The corresponding median of the LTLA-specific correlations between concentration-vaccination interaction and SARS-CoV-2 prevalence is 0.79 across the late period - rounds 12 to 19 (20 May 2021 to 31 March 2022). In further detail, these medians of the LTLA-specific correlations can be interpreted as the average of LTLA-specific relationships between wastewater concentrations and prevalence. The regional and national-level SARS-CoV-2 wastewater concentration (and concentration-vaccination interaction) and REACT-1 prevalence estimates were obtained by population-level weighting. Here, for each period, we measured the correlation between the observed values of such regional and national average concentrations and corresponding prevalence. Note that we defined here (and in our modelling analysis) the concentration-vaccination interaction using the proportion of survey participants that were fully vaccinated (two or more doses) from the previous survey round, and the concentration from the current survey round (due to the apparent time taken for vaccination to impact on subsequent population-level faecal shedding).

Rounds	**Variable**	**Spatial Resolution**	**n**	**r (95% CI)**
3-11	Concentration	LTLA	1572	0.62 (0.59, 0.65)
3-11	Concentration	Regional	81	0.83 (0.75, 0.89)
3-11	Concentration	National	9	0.98 (0.82, 1.00)
12-19	Concentration-Vaccination Interaction	LTLA	2461	0.71 (0.69, 0.73)
12-19	Concentration-Vaccination Interaction	Regional	72	0.89 (0.79, 0.94)
12-19	Concentration-Vaccination Interaction	National	8	0.93 (0.60, 1.00)

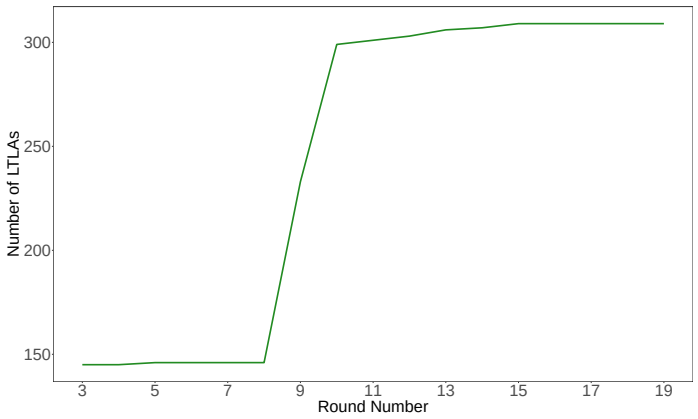


Figure SI A.2: **Number of studied LTLAs by survey round of REACT-1.** The number of LTLA-level observations varied by survey round depending on the number of LTLAs included in the wastewater surveillance programme. Here, we visualise the number of such LTLAs across REACT-1 rounds 3 to 19 (24 July 2020 to 31 March 2022).

A.2 Wastewater-Based Model Covariates

Table SI A.2: **Covariates used in wastewater-based models for estimating SARS-CoV-2 prevalence**. Unless otherwise stated, all covariates are at an LTLA level. Data originate from the REACT-1 study and the EMHP surveillance programme, aside from the LTLA population estimates (provided by the ONS) which are used as weights for regional and national predictors.

Variable	**Units**	**Description**
Wastewater Concentration	Log Gene copies per litre (gc/l)	Estimated LTLA-level wastewater viral concentrations per round. Concentrations are obtained via the weighted contribution of each treatment plant's average (normalised) concentration for the round (described in Materials and Methods).
Neighbour-Averaged Concentration	Log Gene copies per litre (gc/l)	A spatial weights matrix (with row standardisation) is computed based on the neighbourhood structure of LTLAs, using queen contiguity criterion. The weights of this sparse matrix are multiplied by the LTLAs' concentrations for each round to yield a weighted average of the neighbouring areas' concentrations.
Vaccination-Log Concentration Interaction	Log Gene copies per litre (gc/l)	(Fully Vaccinated Proportion) $\times$ (Estimated Concentration). Fully vaccinated proportions are the proportion of the LTLA population estimated (by the REACT-1 study) to have received two or more vaccination doses, whilst concentrations are the estimated LTLA-level concentration.
Regional Average Concentration	Log Gene copies per litre (gc/l)	Average of a region's wastewater concentrations for a round, weighted by the underlying LTLA populations (for whom measurements were available in that round).
National Average Concentration	Log Gene copies per litre (gc/l)	National average of wastewater concentrations for a round, weighted by the underlying LTLA populations (for whom measurements were available in that round).
Concentration Difference	Log Gene copies per litre (gc/l)	The change in an LTLA's wastewater concentration from one round to the next.
National Concentration Difference	Log Gene copies per litre (gc/l)	The change in the national average wastewater concentration from one round to the next.
Historical Regional Prevalence per Log Concentration	% per log gc/l	For each LTLA, prevalence per log concentration is calculated, and the corresponding regional average is the weighted average of the LTLA-level values for that round (where the weights are the LTLA populations). The regional average across the training period is used as the predictor for the testing set. Region-level prevalence per log concentration is used due to the noisiness of the LTLA-level estimates.

A.3 Out-of-Sample Wastewater-Model-Based Estimates of Prevalence

The following section presents the tables and figures which complement the results (see main text) of using a wastewater-based model to estimate SARS-CoV-2 for individual survey rounds which are out-of-sample in the sense that the prevalence data is excluded for all LTLAs in that testing round (and all subsequent survey rounds).

Table SI A.3: **Within-round predictive performance of iteratively-updating wastewater-based model for REACT-1 rounds 7-11 (13 November 2020 to 3 May 2021)**. Summary of accuracy metrics for wastewater-based model's out-of-sample prevalence estimates, where the model was trained iteratively in the sense that the model training window expands over time, to add one additional training round, for each individual testing round. Estimates were out-of-sample in the sense that the prevalence data was omitted for the individual testing round (and all subsequent survey rounds). The intuition is to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey. The wastewater-based (gradient boosting) model was trained at an LTLA level, and regional-level estimates were obtained by weighting the LTLA-level estimates by their corresponding populations within the regions. Here, accuracy metrics have the interpretation that they measure the within-round predictive performance across all n observations, where n is either the number of LTLAs covered by the wastewater surveillance programme or is equal to nine (the number of regions). *Top 25 Common* is the number of LTLAs common to the highest predicted prevalence levels and REACT-1 prevalence levels. *Change Detection* indicates the mean directional accuracy, whilst the corresponding 95% CI is attained by Clopper-Pearson method, otherwise known as the Exact Confidence Interval[482]. *MAE* represents the mean absolute error between each round's model-based predictions and the corresponding LTLA-level (or regional population-weighted average) REACT-1 prevalence estimates. *Mean Prevalence* is the mean average of the individual rounds' LTLA-level (regional population-weighted) REACT-1 prevalence estimates, and is cited as a guide for appraisal of MAE within rounds. Results at a regional level are visualised in Figure SI A.3.

Testing Round	Training Rounds	Resolution	n	Top 25 Common	Change Detection (95% CI)	MAE	Mean Prevalence
7	3-6	LTLA	146	4	71.9% (63.9%,79.0%)	0.57%	1.00%
8	3-7	LTLA	146	3	72.6% (64.6%,79.7%)	0.9%	1.82%
9	3-8	LTLA	233	8	83.7% (78.3%,88.2%)	0.41%	0.49%
10	3-9	LTLA	299	8	72.6% (67.1%,77.6%)	0.23%	0.19%
11	3-10	LTLA	301	3	59.8% (54.0%,65.4%)	0.13%	0.11%
7	3-6	Regional	9		77.8% (40.0%,97.2%)	0.28%	0.97%
8	3-7	Regional	9		88.9% (51.8%,99.7%)	0.56%	1.50%
9	3-8	Regional	9		88.9% (51.8%,99.7%)	0.31%	0.52%
10	3-9	Regional	9		100.0% (66.4%,100.0%)	0.13%	0.22%
11	3-10	Regional	9		88.9% (51.8%,99.7%)	0.08%	0.10%

Table SI A.4: **Predictive performance of iteratively-updating wastewater-based model over time for REACT-1 rounds 7-11 (13 November 2020 to 3 May 2021)**. Summary of accuracy metrics, over five rounds, for a wastewater-based model, which was trained iteratively - in the sense that the model training window expands over time, to add one additional training round, for each individual testing round. Estimates were out-of-sample in the sense that the prevalence data was omitted for each individual testing round (and all subsequent survey rounds). The intuition was to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey. The wastewater-based (gradient boosting) model was trained at an LTLA level, and regional-level estimates were obtained by weighting the LTLA-level estimates by their corresponding populations. Here, accuracy metrics have the interpretation that they measure the predictive performance over time (i.e. five rounds) across all $n^{(\text{obs})}$ observations. *Change Detection* indicates the mean directional accuracy, whilst the corresponding 95% CI is attained by Clopper-Pearson method, otherwise known as the Exact Confidence Interval[482]. r denotes the Pearson's correlation between all wastewater-model-based estimates and REACT-1 SARS-CoV-2 prevalence. *MAE* represents the mean absolute error between all model-based predictions and the corresponding LTLA-level (or regional population-weighted average) REACT-1 prevalence estimates. *Mean Prevalence* is the mean average of the LTLA-level (regional population-weighted) REACT-1 prevalence estimates, and is cited as a guide for appraisal of MAE across the rounds. $n^{(\text{geo})}$ denotes the number of geographies covered by both the wastewater surveillance and community prevalence surveys. Median r (Q1, Q3) denotes the median of the LTLA-level/regional Pearson's correlations between their wastewater-model-based estimates and their corresponding REACT-1 SARS-CoV-2 prevalence. This summary statistic may be useful for understanding how well our wastewater model could track the prevalence trends of individual LTLAs (or regions) for these short periods of time of up to month (and therefore, potentially in between rounds of a community prevalence survey). Results at a regional level are visualised in Figure SI A.3.

Testing Rounds	Training Rounds	Resolution	$n^{(\text{obs})}$	Change Detection (95% CI)	r	MAE	Mean Prevalence	$n^{(\text{geo})}$	Median r (Q1,Q3)
7-11	3-10	LTLA	1125	71.4% (68.6%,74.0%)	0.66	0.37%	0.54%	299	0.85 (0.62, 0.96)
7-11	3-10	Regional	45	88.9% (75.9%,96.3%)	0.79	0.27%	0.66%	9	0.94 (0.85, 0.95)

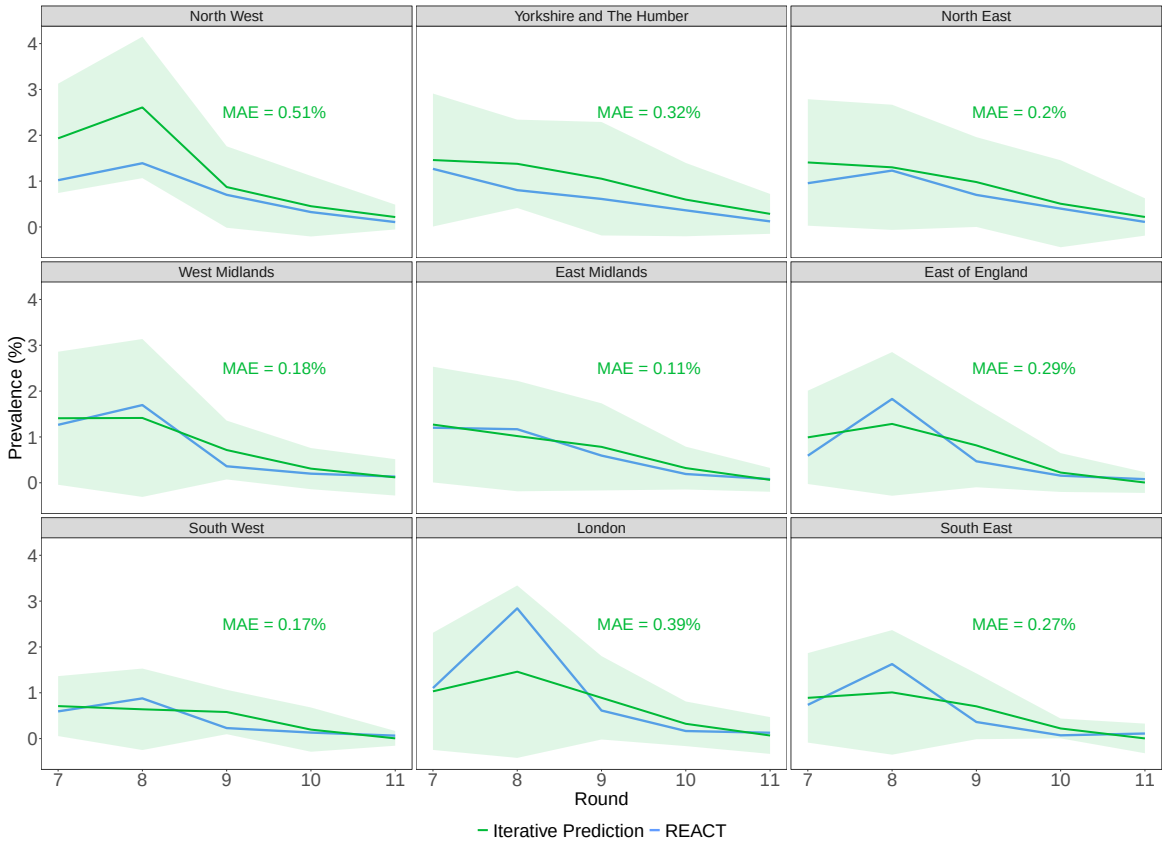


Figure SI A.3: **Regional out-of-sample, wastewater-model-based prevalence estimates in rounds 7 to 11 (13 November 2020 to 3 May 2021).** Wastewater-model-based prevalence estimates are shown alongside REACT-1 prevalence (blue). Estimates were out-of-sample in the sense that the prevalence data was omitted for each individual testing round (and all subsequent survey rounds). The model was trained at an LTLA level with iterative (green) updating of the model (thus expanding the training set every round with an additional training round). The scenario amounts to employing a wastewater-based model for a single survey round (i.e. a relatively short period of time up to one month), and thus assesses predictive performance with regular model calibration of the wastewater-to-prevalence relationship. Such a scenario may arise in between rounds of a community prevalence survey and thus, captures the ability of wastewater-based models for filling temporal gaps in prevalence surveillance. The wastewater-model-based and REACT-1 prevalence estimates for each region (per individual round) were obtained by weighting the LTLA-level prevalence estimates by the corresponding LTLA populations within the region. The shaded 95% prediction intervals (PIs) were obtained by 5,000 non-parametric bootstrap samples. The mean average of the (population-weighted) regional prevalence values across rounds 7 to 11 was 0.66%.

Table SI A.5: **Comparing predictive performance, over time, of an iteratively-updating wastewater-based model's estimates and a wastewater-based model making multi-step prevalence estimates for REACT-1 rounds 9-11 (4 February 2021 to 3 May 2021)**. The wastewater-based model which was trained iteratively meant that the model training window expands over time, to add one additional training round, for each individual testing round. The intuition is to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey. On the other hand, the wastewater-based model which made multi-step prevalence estimates had a fixed training set (of rounds 3 to 7) across the three testing rounds and wastewater-model-based prevalence estimates were made for the individual testing rounds (without any updating of the training set). In each case, the wastewater-model-based estimates for each individual round were out-of-sample in the sense that prevalence data was omitted for that individual round (and all subsequent rounds). The wastewater-based (gradient boosting) models were trained at an LTLA level, and regional-level estimates were obtained by weighting the LTLA-level estimates by their corresponding populations. Here, accuracy metrics have the interpretation that they measure the predictive performance over time (i.e. three rounds) across all $n^{(\text{obs})}$ observations. *Change Detection* indicates the mean directional accuracy, whilst the corresponding 95% CI is attained by Clopper-Pearson method, otherwise known as the Exact Confidence Interval[482]. r denotes the Pearson's correlation between all wastewater-model-based estimates and REACT-1 SARS-CoV-2 prevalence. *MAE* represents the mean absolute error between all model-based predictions and the corresponding LTLA-level (or regional population-weighted average) REACT-1 prevalence estimates. *Mean Prevalence* is the mean average of the LTLA-level (regional population-weighted) REACT-1 prevalence estimates, and is cited as a guide for appraisal of MAE across the rounds. $n^{(\text{geo})}$ denotes the number of geographies covered by both the wastewater surveillance and community prevalence surveys. Median r (Q1, Q3) denotes the median of the LTLA-level/regional Pearson's correlations between their wastewater-model-based estimates and their corresponding REACT-1 SARS-CoV-2 prevalence. This summary statistic may be useful for understanding how well our wastewater model could track the prevalence trends of individual LTLAs (or regions) for either short periods of time of up to month (and therefore, potentially in between rounds of a community prevalence survey) or possibly, longer periods of up to three months. Results for both modelling environments are visualised at a regional level are visualised in Figure 2.

Type	Test Rounds	Training Rounds	Resolution	$n^{(\text{obs})}$	Change Detection (95% CI)	r	MAE	Mean Prevalence	$n^{(\text{geo})}$	Median r (Q1,Q3)
Iterative	9, 10, 11	3-8, 3-9, 3-10	LTLA	1125	71.1% (67.9%,74.1%)	0.54	0.25%	0.24%	299	0.91 (0.41, 0.99)
Multi-Step	9, 10, 11	3-8	LTLA	1125	69.0% (65.8%,72.2%)	0.52	0.27%	0.24%	299	0.87 (0.38, 0.98)
Iterative	9, 10, 11	3-8, 3-9, 3-10	Regional	27	92.6% (75.7%,99.1%)	0.94	0.17%	0.28%	9	1 (0.99, 1)
Multi-Step	9, 10, 11	3-8	Regional	27	77.8% (57.7%,91.4%)	0.93	0.21%	0.28%	9	0.98 (0.96, 0.99)

Table SI A.6: **Within-round predictive performance of iteratively-updating wastewater-based model for REACT-1 rounds 15-19 (19 October 2021 to 31 March 2022)**. Summary of accuracy metrics for wastewater-based model's out-of-sample prevalence estimates, where the model was trained iteratively in the sense that the model training window expands over time, to add one additional training round, for each individual testing round. Estimates were out-of-sample in the sense that the prevalence data was omitted for the individual testing round (and all subsequent survey rounds). The intuition was to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey. The wastewater-based (gradient boosting) model was trained at an LTLA level, and regional-level estimates were obtained by weighting the LTLA-level estimates by their corresponding populations within the regions. Here, accuracy metrics have the interpretation that they measure the within-round predictive performance across all n observations, where n is either the number of LTLAs covered by the wastewater surveillance programme or is equal to nine (the number of regions). *Top 25 Common* is the number of LTLAs common to the highest predicted prevalence levels and REACT-1 prevalence levels. *Change Detection* indicates the mean directional accuracy, whilst the corresponding 95% CI is attained by Clopper-Pearson method, otherwise known as the Exact Confidence Interval[482]. *MAE* represents the mean absolute error between each round's model-based predictions and the corresponding LTLA-level (or regional population-weighted average) REACT-1 prevalence estimates. *Mean Prevalence* is the mean average of the individual rounds' LTLA-level (regional population-weighted) REACT-1 prevalence estimates, and is cited as a guide for appraisal of MAE within rounds. Results at a regional level are visualised in Figure SI A.4.

Testing Round	Training Rounds	Resolution	n	Top 25 Common	Change Detection (95% CI)	MAE	Mean Prevalence
15	3-14	LTLA	309	1	75.1% (69.9%,79.8%)	0.83%	1.58%
16	3-15	LTLA	309	1	74.1% (68.8%,78.9%)	0.79%	1.37%
17	3-16	LTLA	309	2	62.8% (57.1%,68.2%)	2.76%	4.18%
18	3-17	LTLA	309	8	73.8% (68.5%,78.6%)	1.58%	2.84%
19	3-18	LTLA	309	1	91.3% (87.5%,94.2%)	2.31%	6.47%
15	3-14	Regional	9		66.7% (29.9%,92.5%)	0.58%	1.59%
16	3-15	Regional	9		66.7% (29.9%,92.5%)	0.33%	1.37%
17	3-16	Regional	9		66.7% (29.9%,92.5%)	3.17%	4.66%
18	3-17	Regional	9		77.8% (40.0%,97.2%)	1.44%	2.79%
19	3-18	Regional	9		100.0% (66.4%,100.0%)	1.87%	6.26%

Table SI A.7: **Predictive performance of iteratively-updating wastewater-based model over time for REACT-1 rounds 15-19 (19 October 2021 to 31 March 2022)**. Summary of accuracy metrics for wastewater-based model, which was trained iteratively - in the sense that the model training window expands over time, to add one additional training round, for each individual testing round. The intuition was to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey. Estimates were out-of-sample in the sense that the prevalence data was omitted for the individual testing round (and all subsequent survey rounds). The wastewater-based (gradient boosting) model was trained at an LTLA level, and regional-level estimates were obtained by weighting the LTLA-level estimates by their corresponding populations. Here, accuracy metrics have the interpretation that they measure the predictive performance over time (i.e. five rounds) across all $n^{(\text{obs})}$ observations. *Change Detection* indicates the mean directional accuracy, whilst the corresponding 95% CI is attained by Clopper-Pearson method, otherwise known as the Exact Confidence Interval[482]. r denotes the Pearson's correlation between all wastewater-model-based estimates and REACT-1 SARS-CoV-2 prevalence. *MAE* represents the mean absolute error between all model-based predictions and the corresponding LTLA-level (or regional population-weighted average) REACT-1 prevalence estimates. *Mean Prevalence* is the mean average of the LTLA-level (regional population-weighted) REACT-1 prevalence estimates, and is cited as a guide for appraisal of MAE across the rounds. $n^{(\text{geo})}$ denotes the number of geographies covered by both the wastewater surveillance and community prevalence surveys. Median r (Q1, Q3) denotes the median of the LTLA-level/regional Pearson's correlations between their wastewater-model-based estimates and their corresponding REACT-1 SARS-CoV-2 prevalence. This summary statistic may be useful for understanding how well our wastewater model could track the prevalence trends of individual LTLAs (or regions) for these short periods of time of up to month (and therefore, potentially in between rounds of a community prevalence survey). Results at a regional level are visualised in Figure SI A.4.

Testing Rounds	Training Rounds	Resolution	$n^{(\text{obs})}$	Change Detection (95% CI)	r	MAE	Mean Prevalence	$n^{(\text{geo})}$	Median r (Q1,Q3)
15, 16, 17, 18, 19	3-14, 3-15, 3-16, 3-17, 3-18	LTLA	1545	75.4% (73.2%,77.5%)	0.65	1.65%	3.29%	309.00	0.78 (0.57, 0.90)
15, 16, 17, 18, 19	3-14, 3-15, 3-16, 3-17, 3-18	Regional	45	75.6% (60.5%,87.1%)	0.74	1.48%	3.33%	9.00	0.75 (0.65, 0.88)

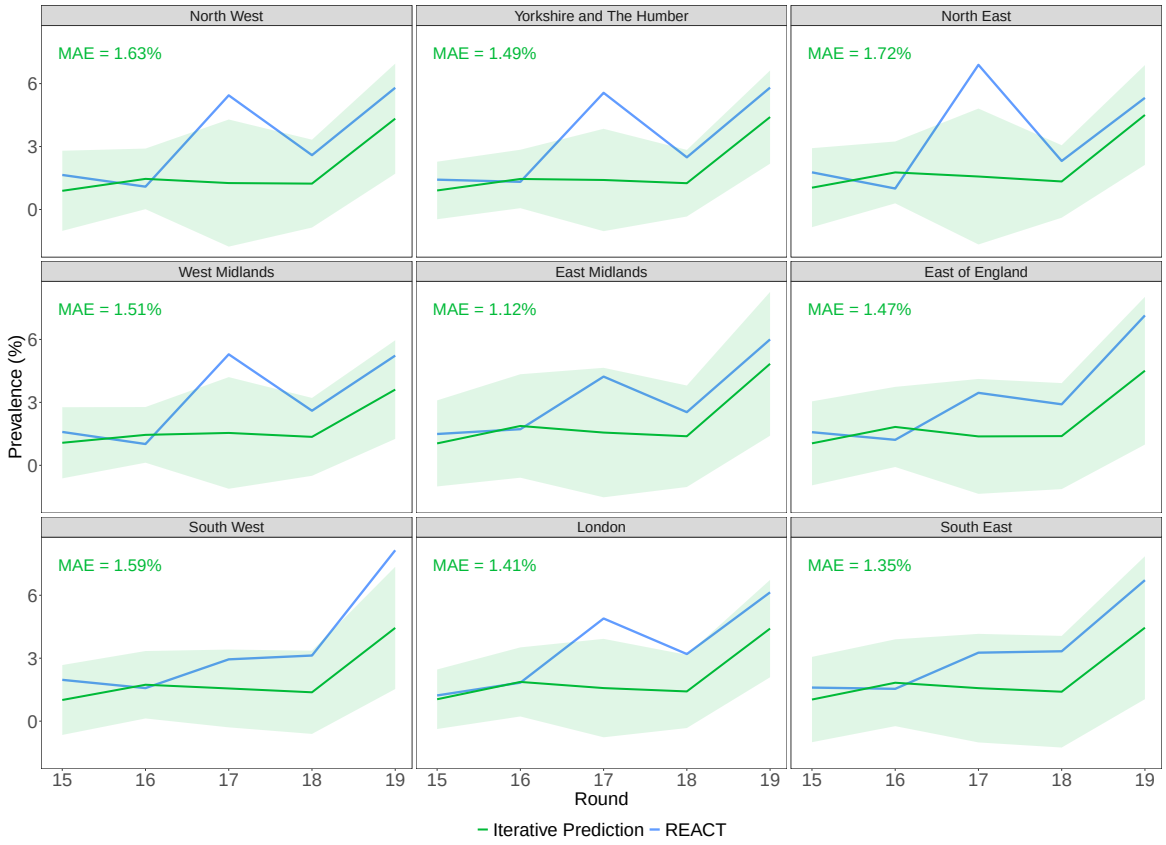


Figure SI A.4: **Regional out-of-sample, wastewater-model-based prevalence estimates in rounds 17 to 19 (19 October 2021 to 31 March 2022).** Wastewater-model-based prevalence estimates are shown alongside REACT-1 prevalence (blue). The model was trained at an LTLA level with iterative (green) updating of the model (thus expanding the training set every round with an additional training round). The scenario amounts to employing a wastewater-based model for a single survey round (i.e. a relatively short period of time up to one month), and thus assesses predictive performance with regular model calibration of the wastewater-to-prevalence relationship. Such a scenario may arise in between rounds of a community prevalence survey and thus, captures the ability of wastewater-based models for filling temporal gaps in prevalence surveillance. The wastewater-model-based and REACT-1 prevalence estimates for each region (per individual round) were obtained by weighting the LTLA-level prevalence estimates by the corresponding LTLA populations within the region. The shaded 95% prediction intervals (PIs) were obtained by 5,000 non-parametric bootstrap samples. The mean average of the (population-weighted) regional prevalence values across rounds 15 to 19 was 3.33%.

Table SI A.8: **Comparing predictive performance, over time, of iteratively-updating wastewater-based model and wastewater-based model making multi-step prevalence estimates for REACT-1 rounds 17 to 19 (19 October 2021 to 31 March 2022)**. The wastewater-based model which was trained iteratively meant that the model training window expands over time, to add one additional training round, for each individual testing round. The intuition was to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey. On the other hand, the wastewater-based model which made multi-step prevalence estimates had a fixed training set (of rounds 3 to 16) across the three testing rounds and wastewater-model-based prevalence estimates were made for the individual rounds (without any updating of the training set). In each case, the wastewater-model-based estimates for each individual round were out-of-sample in the sense that prevalence data was omitted for that individual round (and all subsequent rounds). The wastewater-based (gradient boosting) models were trained at an LTLA level, and regional-level estimates were obtained by weighting the LTLA-level estimates by their corresponding populations. Here, accuracy metrics have the interpretation that they measure the predictive performance over time (i.e. three rounds) across all $n^{(\text{obs})}$ observations. *Change Detection* indicates the mean directional accuracy, whilst the corresponding 95% CI is attained by Clopper-Pearson method, otherwise known as the Exact Confidence Interval[482]. r denotes the Pearson's correlation between all wastewater-model-based estimates and REACT-1 SARS-CoV-2 prevalence. *MAE* represents the mean absolute error between all model-based predictions and the corresponding LTLA-level (or regional population-weighted average) REACT-1 prevalence estimates. *Mean Prevalence* is the mean average of the LTLA-level (regional population-weighted) REACT-1 prevalence estimates, and is cited as a guide for appraisal of MAE across the rounds. $n^{(\text{geo})}$ denotes the number of geographies covered by both the wastewater surveillance and community prevalence surveys. Median r (Q1, Q3) denotes the median of the LTLA-level/regional Pearson's correlations between their wastewater-model-based estimates and their corresponding REACT-1 SARS-CoV-2 prevalence. This summary statistic may be useful for understanding how well our wastewater model could track the prevalence trends of individual LTLAs (or regions) for these short periods of time of up to month (and therefore, potentially in between rounds of a community prevalence survey). Results for both modelling environments are visualised at a regional level are visualised in Figure 2.

Type	Test Rounds	Training Rounds	Resolution	$n^{(\text{obs})}$	Change Detection (95% CI)	r	MAE	Mean Prevalence	$n^{(\text{geo})}$	Median r (Q1,Q3)
Iterative	17, 18, 19	3-16, 3-17, 3-18	LTLA	1545	75.9% (73.1%,78.7%)	0.60	2.22%	4.5%	309	0.89 (0.53, 0.98)
Multi-Step	17, 18, 19	3-16	LTLA	1545	52.0% (48.7%,55.3%)	0.23	3.08%	4.5%	309	0.62 (0.08, 0.91)
Iterative	17, 18, 19	3-16, 3-17, 3-18	Regional	27	81.5% (61.9%,93.7%)	0.74	1.48%	4.57%	9	0.85 (0.59, 0.99)
Multi-Step	17, 18, 19	3-16	Regional	27	48.1% (28.7%,68.1%)	0.39	3.09%	4.57%	9	0.88 (0.57, 0.93)

A.4 Complementary Use of WBE

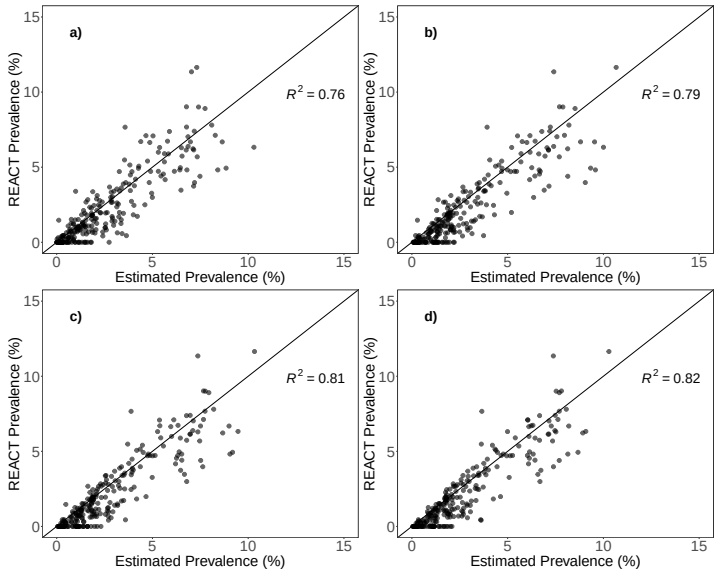


Figure SI A.5: **Wastewater-model-based prevalence estimates using varying training set sizes.** From plots a) to d), across rounds 12 to 19 (from 20 May 2021 to 31 March 2022), training set sizes are 40%, 60%, 80%, and 90% respectively per each round's total number of LTLAs. The plot above is for a randomly-selected fold amongst the 50 folds considered in our replicated procedure for training-testing split proportions, and wastewater-model-based prevalence estimates of a fixed test set improve marginally as more training observations are used to calibrate the model. Nevertheless, with just 40% survey coverage, wastewater-model-based estimates remain largely representative of underlying prevalence.

Table SI A.9: **Comparison of wastewater-model-based prevalence estimates in the *early* period of rounds 3 to 11 (24 July 2020 to 3 May 2021) using varying training test set sizes**. Wastewater-based gradient boosting models are trained using 40%-90% of each round's observations and a fixed 10% of observations per round are used for test set prevalence estimation. The average prevalence across the 50 folds of test sets is 0.54%.

Training-Testing	MAE	r	Change Detection
40% - 10%	0.34 %	0.74	74.57%
60% - 10%	0.33 %	0.74	75.14%
80% - 10%	0.32 %	0.76	76.30%
90% - 10%	0.32 %	0.75	74.57%

A.5 Population-level faecal shedding

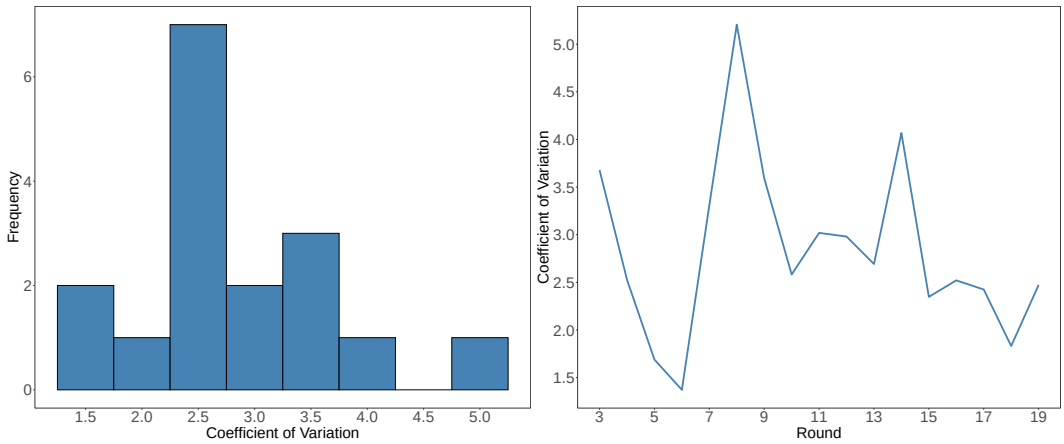


Figure SI A.6: **Spatial variability of faecal shedding by round.** Visualisations demonstrate the spatial consistency of the prevalence-to-wastewater relationship. The histogram (left) of the coefficient of variation (CV) values demonstrate that extreme large values of the measure of dispersion are not common, whilst the line plot (right) displays the spatial consistency over time.

A.6 Geospatial Population Estimates (GPEs)

The following section provides further detail regarding the derivation and uncertainty of the GPEs used for our geospatial framework and associated analyses.

Let a_{ij} be the intersection area between the catchment area of STW i and LSOA j. Then, the LSOA area serviced by any of the STWs i is:

$$A_j = \bigcup_i a_{ij}$$

Let p_j be the Office for National Statistics (ONS) 2019 mid-year population estimate for LSOA j. Our geospatial population estimate (GPE) for the intersection area between the catchment area of STW i and the area of LSOA j is

$$\widehat{g_{ij}} = \frac{a_{ij}}{A_j} p_j$$

Thus, for LSOA j, our estimate of the LSOA population serviced is:

$$\widehat{p_j} = \sum_i \widehat{g_{ij}}$$

Summing over all LSOA geographies within an LTLA k, we attain our GPE $\widehat{P_k}$ for each LTLA:

$$\widehat{P_k} = \sum_{j=1}^{J} \widehat{p_j}$$

The GPE for each LTLA enables our mapping of concentrations from an STW level to an LTLA level.

Our approach of using GPEs represents a predictive approach for aligning WBE with the geographies commonly used by public health authorities[493]. Consistent with the reported EMHP population coverage estimates[302], we estimate that the wastewater programme had an estimated 74% nationwide testing coverage by its conclusion in March 2022, with a median percentage (of population) sampled for LTLAs of 76%.

The developed approach could, in theory, be applicable to further population-level wastewater monitoring and analyses which require alignment of wastewater catchments to geographies used by public health authorities and community studies. Nevertheless, limitations exist with our geospatial approach. First, the GPEs do not account for the presence of transient and non-resident LTLA populations, which could impact on the wastewater concentrations measured. The transience issue is likely to be more pronounced in the late rounds of the REACT-1 analysis, where community lockdowns and other non-pharmaceutical interventions are relaxed. Second, we recognise the uncertainty surrounding the mid-year population estimates reported by the ONS for LSOAs, which are made for small geographic areas.

Furthermore, potential limitations may yield unrepresentative population estimates for urban centers and/or for geographies which are influenced by time-varying factors such as transient/commuting populations or industrial and agricultural discharges. In theory, such factors are controlled via our flow-normalised wastewater concentrations yet cannot be taken into account when we weight the contribution of each STW to an LTLA based on time-invariant population estimates). The issue of time-varying relationships is likely to depend upon the type and stringency level of simultaneously active non-pharmaceutical interventions.

Further challenges imposed by the EMHP wastewater surveillance programme include the locations of the 302 STWs. Whilst the sampled treatment plants were selected to maximise nationwide coverage and representativeness across England, our GPEs (for each LTLA) indicate a highly positively skewed distribution for the estimated proportions of individual LTLA populations that are sampled within the EMHP programme (Figure SI A.7). The skewness in the distribution of estimated proportions of populations sampled is a likely consequence of the usage of intersection areas in our GPEs. We estimate the median LTLA-level proportion of population sampled to be approximately 76%, yet Horesham LTLA, for instance,

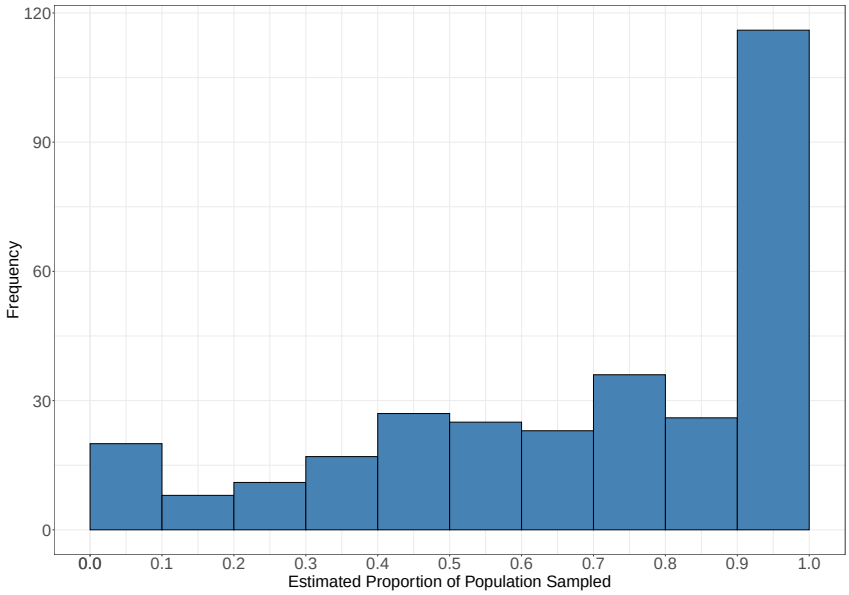


Figure SI A.7: **LTLA-level wastewater sampling coverage estimates.** The estimated proportions of LTLA populations that are covered by the EMHP wastewater surveillance programme. The proportions are attained by dividing the LTLA-level GPE by the corresponding ONS LTLA population approximation.

has an estimated testing coverage of only 2.9% based on the GPE, consistent with the EMHP testing coverage estimate for Horesham. These few outliers can potentially impact on how representative our wastewater concentrations are when mapped to LTLAs with such low testing coverage, and hence, when relating our wastewater data to LTLA-level survey-based prevalence estimates.

A.7 Further Details of REACT-1

The REACT-1 study was described briefly in Materials and methods. Here, we outline further details of the study's design and statistical analyses.

Study design

Throughout the 19 distinct rounds, individuals were sampled from the National Health Service (NHS) general practitioner list of patients in England. The list provides details such as name, month and year of birth, gender, and address. To achieve the study's targeted within-round sample size (of between 100,000 and 150,000), up to 750,000 individuals above the age of five years old were random sampled from the general practitioner list and the study team (researchers from Imperial College London and its logistics partner Ipsos) sent personalised invitations to individuals (by mail). Potential participants were provided with study details and asked to indicate their willingness to participate (either via an online portal or via telephone). The participants were provided with both written and video instructions for completing the throat-and-nose swab. If participants provided a swab, they then completed a detailed questionnaire which included questions such as household composition, contact with people known or suspected to have COVID-19, and vaccination history [306,307,308].

In terms of the sampling strategy, from round 1 (1 May 2020 to 1 June 2020) to round 12 (20 May 2021 to 7 June 2021), the objective was to obtain LTLA-stratified random samples with approximately equal numbers of participants in each of the 315 LTLAs. However, from round 12 (20 May 2021 to 7 June 2021) to round 19 (8 March 2022 to 31 March 2022), the sampling procedure was adjusted to sample randomly in proportion to LTLA-level populations and thus, correct for under-representation in urban and inner-city areas (relative to more rural and sparsely populated areas) [308].

From round 1 (1 May 2020 to 1 June 2020) to round 13 (24 June 2021 to 12 July 2021), dry swab samples were delivered to the laboratory by a courier on a cold chain. In round 14 (9 September 2021 to 27 September 2021), the study changed to usage of wet (saline) swabs which were sent to the laboratory either by courier (without a cold chain) or priority post. From round 15 (19 October 2021 to 5 November 2021) to round 19 (8 March 2022 to 31 March 2022), the swabs were sent by priority post only. Due to postal-related delays for swabs, the study included a small number of samples obtained after the specified closing date for rounds 14 to 18 [306,307,308].

Statistical analysis

The random iterative method (RIM) weighting aimed to correct for the non-response bias in (unweighted) prevalence estimates [492]. In particular, to produce the LTLA-level weighted prevalence estimates, the analysis considered varying non-response in varying demographic groups/categories, which included age-sex categories, deprivation index deciles, ethnicity, and LTLA-level populations.

The RIM weighting involved two steps. In the first step, each sample was weighted to LTLA counts and age-sex categories, and the second step adjusted the RIM weighting for all four of the weighting variables. Extreme weights were dampened by trimming the adjustment factor between the first and second step weights at the 1st and 99th percentiles. Thus, the final weights were the first-stage weights multiplied by the trimmed adjustment factor for the second stage [308].

A.8 Wastewater Sources of Uncertainty

Sources of wastewater measurement uncertainty include the sample volume being too low to enable adequate analysis, temperature and time-induced decay, inherent variability of wastewater (possibly due to dilution effects of precipitation), and usage of the Theoretical Limit of Detection (TLoD), below which the concentration cannot be reliably measured. The EMHP

surveillance programme aimed to address several of these measurement uncertainties via the adjustment/normalisation of concentrations to account for flow, and by sampling mid-stream during peak load times. Similarly, 3 to 4 samples were taken weekly (at each STW) due to the variability and presence of outliers in detected wastewater signals[302,489].

A.9 Sensitivity of Geospatial Mapping Approach

Table SI A.10: **Sensitivity analysis for geospatial mapping**. Table depicts a diminished relationship between wastewater concentrations and SARS-CoV-2 prevalence, when concentrations from rounds of the REACT-1 are shifted by a lead times of up to six days. Overall correlation measures the correlation between LTLA-level estimated wastewater concentrations and SARS-CoV-2 prevalence. Mean LTLA concentration correlation is the mean of all the LTLA's individual correlations between estimated wastewater concentrations and SARS-CoV-2, whilst the mean LTLA interaction correlation (applicable for rounds 12 to 19) is the average of the LTLA's individual correlation between our proposed vaccination-log concentration interaction and SARS-CoV-2 prevalence.

Lead Time (Days)	Overall Correlation	Mean LTLA Concentration Correlation	Mean LTLA Interaction Concentration Correlation
0	0.35	0.39	0.75
1	0.32	0.33	0.75
2	0.29	0.31	0.75
3	0.28	0.30	0.75
4	0.28	0.25	0.74
5	0.27	0.25	0.74
6	0.19	0.16	0.66

A.10 Bayesian Hierarchical Modelling

To ensure robustness of inferences and extend our analysis beyond a discriminative model, within the Bayesian modelling paradigm, we fitted a hierarchical model, which is a generative model type which enables regularisation and spatial effects. Our hierarchical model is fit to training periods using Stan v2.31, which employs Hamiltonian Monte Carlo sampling, a variant of Markov Chain Monte Carlo (MCMC)[763].

Let y_{it} be the REACT-1 weighted prevalence estimate for LTLA i at time/round t. Then, our varying effects Bayesian hierarchical model is as follows:

$$
\begin{aligned}
y_{it} &\sim \text{Normal}(\mu_{it}, \sigma^2) \\
\mu_{it} &= \beta_0 + \beta_{LTLA[i]} * C_{it} + \beta_{R[i]} * C_{it} + \beta_{NN[i]} * NC_{it} \\
\beta_0 &\sim \text{Normal}(0, 5) \\
\beta_{LTLA} &\sim \text{Normal}(0, \sigma^2_{LTLA}) \\
\sigma_{LTLA} &\sim \text{InvGamma}(5, 5) \\
\beta_R &\sim \text{Normal}(0, \sigma^2_R) \\
\sigma_R &\sim \text{InvGamma}(5, 5) \\
\beta_{NN} &\sim \text{Normal}(0, \sigma_{NN}) \\
\sigma_{NN} &\sim \text{InvGamma}(5, 5) \\
\sigma &\sim \text{Half-Cauchy}(0, 1)
\end{aligned}
$$

where C_{it} denotes the concentration of LTLA i at time/round t, and NC_{it} denotes the weighted average concentration for the neighbours of LTLA i at time/round t. μ and σ represent the mean and standard deviation of the assumed Normal distribution. β_0 is the baseline prevalence. β_{LTLA}, β_R, and β_{NN}, are parameter vectors corresponding to the LTLA-specific, region-specific, and nearest-neighbour specific effects of concentration on the response of prevalence. σ_{LTLA}, σ_R, and σ_{NN} represent the corresponding standard deviations. Above, we employ an index-variable approach within the linear predictor. For example $\beta_{R[i]}$ maps the LTLA i to its corresponding parameter within the region-effect concentration vector. The intuition for spatially-varying effects is that in-sewer network characteristics, treatment plant-level, and spatial correlations between geographies could influence similarities and differences of both concentrations and prevalence levels. For the *late* period of the REACT-1 analysis, the vaccination-log concentration interaction (described in Table SI A.2) takes the place of each of the unadjusted wastewater concentrations.

Throughout the analysis, all covariates are standardised such that they take values in $[-1, 1]$. Standardisation facilitates the above usage of conventional, weakly informative prior distributions throughout our model fitting, and assists in ensuring efficient HMC sampling.

Similar to the gradient boosting setup, the model was trained using several calibration rounds from REACT-1, and its predictive performance was estimated out-of-sample using one or more rounds of REACT-1. Our model selection procedure involved consideration of the scientific model's structure, posterior predictive checks, model convergence diagnostics, and estimated out-of-sample predictive accuracy (via the ELPD, expected log pointwise predictive density, from the `loo` package).

In terms of comparing predictive performance to our gradient boosting model, we generate posterior predictions of the response y_{it}. In particular, we use the posterior samples for each observation, and we adopt a conventional approach of using the posterior median of these samples as our best estimate (for each observation). We quantify the uncertainty in our

inferences by deriving credible intervals (by taking specified upper and lower quantiles of these samples). Subsequently, metrics such as MAE, for example, are computed as the average absolute difference between the observed prevalence levels and the posterior median estimates.

Throughout both study periods, our best-performing Bayesian hierarchical model did not provide additional predictive accuracy in terms of out-of-sample predictive accuracy. Superior accuracy of the gradient boosting model may be a consequence of the highly flexible and predictive nature of the modern gradient boosting algorithm (of `xgboost`) which readily handles time-varying non-linearities. Nevertheless, the additional modelling investigation enabled a further robustness check of the inferences drawn regarding the reliability of wastewater-based modelling.

A.11 Vaccination-Log Concentration Interaction Variable

In terms of the statistical controls in our analysis, there exist obvious limitations with our proposed usage of a simple vaccination-concentration interaction. Aside from the potential differential impacts of vaccination on distinct demographies, the proposed interaction variable would only remain (possibly) valid for a restricted time period due to impacts of waning vaccine-induced immunity. We further do not account for the highly complex nature of naturally-acquired immunity. The interaction variable's usage is additionally challenged by the disparate impact of different variants (and sub-variants) on the likelihood of SARS-CoV-2 reinfection. Importantly, we do not draw *any causal conclusions* regarding the impact of vaccination on faecal shedding due to the potential presence of confounding. Indeed, we emphasise, that strong temporal correlation is not necessarily indicative of a direct (or causal) relationship with prevalence, as the directional association may be an artefact of numerous epidemic conditions. Complex, concurrent epidemic characteristics include a monontonically increasing vaccination proportion, similarly rising prevalence levels, waning immunity (either vaccine-induced or naturally-acquired), reduced immunity to particular new variants, and other possibly unobserved confounders.

A.12 Acronyms

Table SI A.11: **Acronyms used throughout the manuscript of Chapter 3**

Variable	Meaning	Details
95% CrI	95% Credible Interval	An unobserved parameter lies in this interval with a specified 95% probability.
EMHP	Environmental Monitoring for Health Protection	The EMHP wastewater surveillance programme tested untreated sewage across England for fragments of SARS-CoV-2. The objectives of the programme were to monitor wastewater viral concentrations of SARS-CoV-2 RNA, variants of concern (VOC), and variants under investigation (VUI).
gc/l	Gene copies per litre (gc/l)	The reported measuresments of wastewater concentrations obtained via RT-qPCR and flow normalisation.
GPE	Geospatial population estimate	Our population estimates for geographies based on a combination of intersection (spatial) areas and reported population estimates for small geographies.
LSOA	Lower Layer Super Output Area	Small regional geographies in England which combine to form an LTLA.
LTLA	Lower Tier Local Authority	LTLA-level wastewater measurements per round. These are obtained via the weighted contribution of each treatment plant's average concentration for the round (described in Materials and Methods).
MAE	Mean Absolute Error	The mean average of the absolute difference between the wastewater-model-based estimates and the REACT-1 prevalence levels.
MCMC	Markov Chain Monte Carlo	A family of sampling algorithms which employs the theory of Markov Chains to sample a random variable.
NHS	National Health Service	The public healthcare system in the United Kingdom.
NPI	Non-pharmaceutical intervention	In the context of the COVID-19 pandemic, these are public health measures, excluding medication-based measures, taken to control transmission of the SARS-CoV-2.
ONS	Office for National Statistics	An independent producer of national statistics across the UK.
PI	Prediction Interval	We use 95% Prediction Intervals for our wastewater-based (gradient boosting) model's prevalence estimates which were estimated by 5,000 non-parametric bootstrap replicates of training datasets.
REACT-1	Real-time Assessment of Community Transmission	The REACT-1 programme was initiated in May 2020 with an objective of tracking the spread of SARS-CoV-2 across communities in England. Across 19 distinct rounds of cross-sectional surveys, random samples of the English population (over 5 years of age) were taken. Rounds lasted between 15 and 31 days, and the programme concluded in March 2022.
RNA	Ribonucleic acid	RNA is a nucleic acid present in all living cells.
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction	The commercial laboratory of the REACT-1 study used RT-PCR. The method combines reverse transcription of RNA into DNA and amplification of specific DNA targets using polymerase chain reaction.
RT-qPCR	Reverse Transcriptase quantitative Polymerase Chain Reaction	The quantification method is described in [294]. Briefly, RT-qPCR combines reverse transcription and quantitative PCR, with the aim of reducing inhibition in RNA.
STW	Sewage treatment works	STWs are treatment plants which typically service extensive urban areas (like towns and cities).
TLoD	Theoretical limit of detection	The wastewater concentration (160 gc/l) below which the EMHP estimate that concentration cannot be reliably estimated.
UKHSA	United Kingdom Health Security Agency	A nationwide organisation in the United Kingdom which assumes responsibility for public health protection.
VOC	Variants of Concern	SARS-CoV-2 variants that were highlighted by the World Health Organisation (WHO) to be particularly dangerous in terms of increased transmissibility.
VUI	Variants under investigation	SARS-CoV-2 variants were being tracked by the EMHP wastewater surveillance programme.
WBE	Wastewater-based epidemiology	WBE involves collection of urine and stool samples from sewage treatment works (STWs). By subsequently incorporating factors such as daily flow rates, human excretion rates, and STW catchment population sizes, per-capita consumption, use, or exposure can be obtained.
WHO	World Health Organisation	An agency of the United Nations which aims to improve international public health.

A.13 Timeline of REACT-1 and Estimated Number of LTLAs in EMHP Programme

Table SI A.12: Dates (DD/MM/YYYY) for individual rounds of the REACT-1 study and the corresponding estimated number of LTLAs which were mapped to STWs which reported measurements within the corresponding time intervals. Our overall studied period, where the EMHP wastewater surveillance programme coincides with the REACT-1 study, covers from rounds 3 to 19 (from 24 July 2020 to 31 March 2022).

Round Number	Round Start Date	Round End Date	Number of LTLAs
1	01/05/2020	01/06/2020	-
2	19/06/2020	07/07/2020	-
3	24/07/2020	11/08/2020	145
4	20/08/2020	08/09/2020	145
5	18/09/2020	05/10/2020	146
6	16/10/2020	02/11/2020	146
7	13/11/2020	03/12/2020	146
8	06/01/2021	22/01/2021	146
9	04/02/2021	23/02/2021	233
10	11/03/2021	30/03/2021	299
11	15/04/2021	03/05/2021	301
12	20/05/2021	07/06/2021	303
13	24/06/2021	12/07/2021	306
14	09/09/2021	27/09/2021	307
15	19/10/2021	05/11/2021	309
16	23/11/2021	14/12/2021	309
17	05/01/2022	20/01/2022	309
18	08/02/2022	01/03/2022	309
19	08/03/2022	31/03/2022	309

A.14 Wastewater Variant Detections

From a public health perspective and for our analysis, the ability to track variants is important to understand the relationship between wastewater concentrations and our community prevalence estimates. Specifically, the EMHP programme raised the conjecture of an Omicron effect which results in lower viral faceal shedding and substantially alters the relationship between wastewater signals and clinical measures[302]. The variant detections were attained via genomic sequencing of wastewater samples from both STWs and sewer network sites across England, and detections were reported as either a confirmed or a possible status. More recently, the importance of accounting for the predominant variant was demonstrated by research of clinical cases showing that Omicron infections yielded the lowest community-level SARS-CoV-2 waste shedding rates, compared to the early/parental SARS-CoV-2 and Delta variant[303]. Within our analysis, we have documented the apparent reduced population-level faecal shedding induced by the onset of the Omicron BA.1 and BA.2 sub-variants.

A.15 Further Results

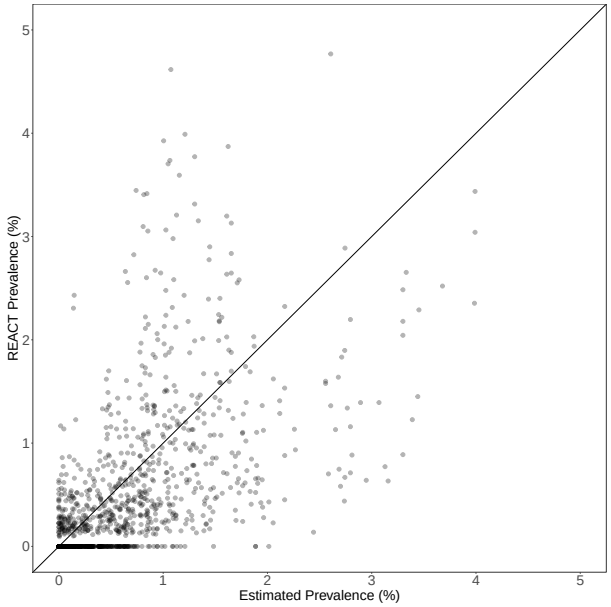


Figure SI A.8: **Iteratively-updating wastewater-model-based prevalence estimates for rounds 7 to 11 (13 November 2020 to 3 May 2021).** Visualisation depicts REACT-1 prevalence versus wastewater-model-based estimates for the corresponding round. The wastewater-based model was trained iteratively - in the sense that the model training window expands over time, to add one additional training round, for each individual testing round. The intuition is to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey.

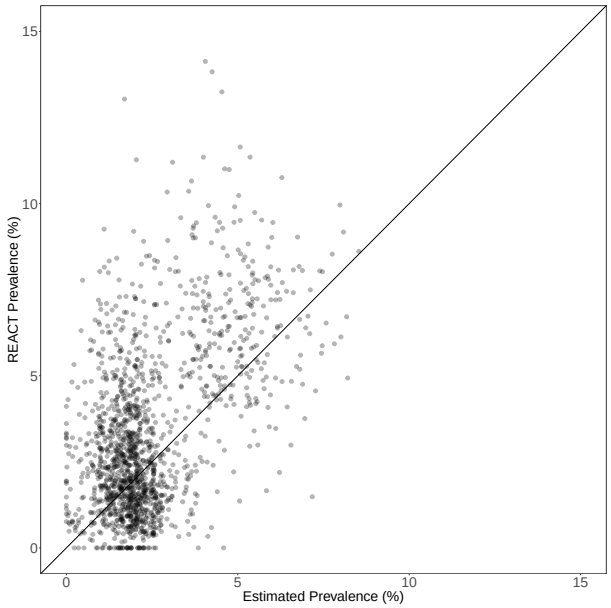


Figure SI A.9: **Iteratively-updating wastewater-model-based prevalence estimates for rounds 15 to 19 (19 October 2021 to 31 March 2022).** Visualisation depicts REACT-1 prevalence versus wastewater-model-based estimates for the corresponding round. The wastewater-based model was trained iteratively - in the sense that the model training window expands over time, to add one additional training round, for each individual testing round. The intuition is to assess how well a wastewater-based model could estimate SARS-CoV-2 prevalence for a relatively short period of time, such as temporarily between rounds of a prevalence survey.

Appendix B

Appendix — Climate-based modelling and forecasting of dengue in three endemic departments of Peru

B.1 Acronyms

Table SI B.1: **Summary of acronyms used throughout the first manuscript of Chapter 4.**

Abbreviation	Definition
Ae. aegypti	*Aedes aegypti*
ACF	Autocorrelation Function
AUC	Area Under the Curve
BYM	Besag, York and Mollié
CV	Cross-Validated
DIC	Deviance Information Criterion
DIR	Dengue Incidence Rate
DLNM	Distributed Lag Nonlinear Models
ENSO	El Niño Southern Oscillation
ICEN	El Niño Coastal Index
IGP	Geophysical Institute of Peru
INEI	National Institute of Statistics and Information of Peru
INLA	Integrated Nested Laplace Approximation
LOOCV	Leave-One-Out-Cross-Validation
MAE	Mean Absolute Error
ONI	Oceanic Niño Index
PIT	Probability Integral Transform
prec	Precipitation
ROC	Receiver Operating Characteristic
RR	Relative Risk
RSI	Relative Strength Index
RW	Random Walk
SPI-6	Standardized Precipitation Index (6-Month)
SST	Sea Surface Temperature
tmax	Maximum Temperature
tmin	Minimum Temperature

B.2 Further study area details

Table SI B.2: **Summary statistics for the studied Peruvian departments.** *Population* refers to the last mid-year population estimate of a department's population within the studied period, as of 30 June 2021, whilst *Population Density* is the corresponding estimated population per km^2. *DIR* is the mean dengue incidence rate per 100,000 across the studied period of 2010 to 2021. *Maximum Temperature* and *Minimum Temperature* similarly relate to the mean average observed value of the monthly means of maximum daily temperatures and of the monthly means of minimum daily temperatures. *Total Precipitation* is the mean of monthly observed total precipitations.

Department	Population	Population Density	DIR	Maximum Temperature	Minimum Temperature	Total Precipitation
Lambayeque	1,325,912	93·0	4·1	27·6°C	16·4°C	13·8 mm
Piura	2,077,039	58·1	31·9	27·8°C	16·1°C	25·4 mm
Tumbes	255,712	55·0	64·6	29·7°C	19·4°C	27·4 mm

B.3 Further model framework details

The following section provides additional details to the description of the modelling framework presented in Section 4.2.2.

B.3.1 Population Offset

For each year and each department, we had access to mid-year population estimates published by the INEI (National Institute of Statistics and Information of Peru) on 30th June of each year[506].

Then, for each month ($t = 1, \ldots 140$), we linearly interpolated between the mid-yearly population values to avoid improbable population fluctuations which would otherwise arise in a single month. In Equation C.4.1, $\log(P_{it})$ meant the natural logarithm of the (interpolated) population of department i at time t divided by 100,000. The population offset enabled our modelling of the logarithm of the dengue incidence rate (DIR) via $\log(\eta_{it})$.

B.3.2 Fixed effects

Our fixed effects, expressed as a matrix of two columns $\boldsymbol{x_i}$ in Equation C.4.1, comprised of; i) our seasonality indicator for the summer months (= 1 for months of December to April, and 0 otherwise) common to each department and ii) the our momentum oscillator; Relative Strength Index (RSI), lagged by one month. The objective of the simple binary seasonality variable was to capture seasonal trends of the summer months which are common across the departments. Whilst our climatic variables accounted for the influence of environmental conditions on vector abundance and incidence rates, the seasonality variable provided a means of accounting for possible population transience and increased outdoors activity during summer months. Meanwhile, the proposed concept of the RSI variable was motivated by its usage as a technical indicator in financial markets to measure the strength and momentum of price movements[518]. For each department, we employed the RSI based on the smoothed moving average of the previous three month's DIRs. Precisely, the formula used for RSI was:

$$\text{RSI} = 100 - \frac{100}{(1 + \text{RS})}$$

where RS is the *relative strength*, defined as the smoothed ratio of average gains (sum of the upward DIR movements over the past three months, divided by three - the number of months) over the average losses (sum of the downward DIR movements over the past three months, divided by three). Here, by using the RSI over the past three months, we aimed to track the recent momentum in DIR movements and provide intuition for future possible DIR movements.

B.3.3 Temporal random effects

Our temporal random effects comprised of: i) a department-specific, cyclic Random Walk (RW) model of order one for the monthly random effects, and ii) department-specific yearly, exchangeable random effects. First, our cyclic RW(1) prior distribution allowed for each department's current month's DIR to depend upon the observed DIR of the previous month, thus introducing temporal autocorrelation into our model structure. The cyclic aspect allowed for the DIR in January of a given year to be related to the DIR observed in the preceding month (December) in the previous year. We also experimented with the usage of a RW(2) prior distribution (a random walk of order two) which would allow for the current DIR observation to depend upon the two previously observed states (two previous month's DIR). Second, our department-specific yearly, exchangeable random effects meant that we adjusted for the possibility that each department may have observed distinct year-to-year heterogeneities such as changes to vector control programmes or varying reporting rates (possibly due to overwhelming of health services or alternatively improvements in epidemiological surveillance). Yearly random effects common across departments were found to be too coarse to enable sufficient description of the year-to-year variability within departments.

B.3.4 Spatiotemporal random effects

Our decision to employ the BYM2 model, an adapted Bayesian version of the Besag-York-Mollié (BYM) model, was motivated by two reasons; i) to adequately capture spatial autocorrelation between the three neighbouring departments, and ii) to account for any non-spatial heterogeneity[519]. In particular, we used a version of the BYM model which employed Penalised Complexity prior distributions (to guard against overfitting) and divided the model into the two components; a spatially structured component and a spatially unstructured component[520]. Spatial autocorrelation means that even after adjusting for several covariates (such as climatic influences), we expected that the case counts (and DIRs) in neighbouring departments would not be wholly independent and would share similarities. The inclusion of non-spatial heterogeneity was useful in terms of adjusting for several of the unobserved confounders (discussed in Section 4.2.4).

In general, the BYM2 model comprises of two hyperparameters; a marginal precision parameter τ and a mixing parameter ϕ. Specifically, our penalised complexity prior on the precision parameter τ meant that $P(\frac{1}{\tau} > 0 \cdot 5) = 0 \cdot 01$. In terms of our model formula (Equation C.4.1), the mixing parameter ϕ controlled the proportion of variation explained by our spatially structured component ($\mathbf{u}$) and our spatially unstructured component ($\mathbf{v}$). The spatially structured component ($\mathbf{u}$) was modelled as a conditional autoregressive process, and represented spatial variation not accounted for by our various other model components. The spatially unstructured component ($\mathbf{v}$) was modelled using an independent and identically distributed Gaussian process.

B.3.5 Specification of climate-based DLNMs

Our model framework employed distributed lag nonlinear models (DLNMs) to capture the potentially nonlinear and delayed effects of climatic variables on dengue incidence rates. DLNMs are a flexible modelling framework which are based on the statistical concept of a *crossbasis*; a bi-dimensional functional space comprised of a combination of two sets of basis functions which represent the relationships with the response in the dimensions of the predictor and lags respectively[394]. In particular, usage of the DLNM framework involves the simultaneous specification of i) nonlinearity in the exposure-response relationship and ii) delayed effects in the lag-response dimension. The specification is achieved via the *cross-basis*, thus creating an exposure-lag-response function.

For our model framework, using the R package `dlnms`, we employed basis splines (B-splines) for our two sets of basis functions, which were combined to generate our cross-basis functions[394]. Specifically, in the exposure-response dimension, we specified B-splines with two equally-spaced knots to enable sufficient flexibility in the modelling of the potentially

complex relationship between the climatic exposure and response. For each climatic variable, the set of basis functions in the exposure-response were centred on the mean observed value of the climatic variable. In the lag-response dimension (up to four months), we allowed for complexity in the delayed effects of climatic variables by specifying a single knot at two months. Sensitivity analyses considered various alternative specifications such as the specification of a single knot a one or three months in lag-response dimension, a simpler specification of a single knot in the exposure-response dimension, and the usage of natural splines (instead of B-splines). We assessed model performance with varying inclusions of the DLNMs for the six climatic variables (which were described in Section 4.2.2 and visualised in Figure SI B.7).

B.4 Figures

B.4.1 Exploratory analyses

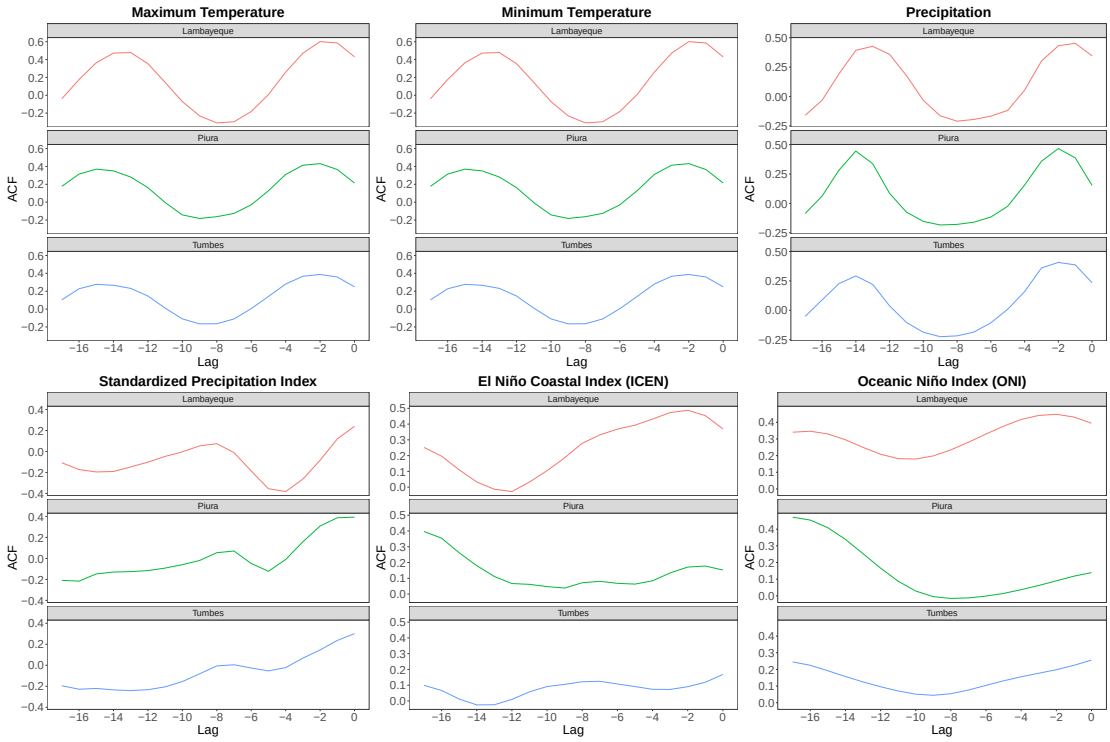


Figure SI B.1: **Exploratory analysis of climate-dengue relationships.** Cross-correlation plots for monthly lags of six climatic variables and the monthly dengue incidence rate (DIR) per 100,000 in each of the three studied departments. The plots depict a strong degree of shared similarity across departments in the timing and nature of relationship between the climate exposure and DIR. Negative lags correspond to lead times in months.

B.4.2 Sensitivity analyses

Climatic inference without seasonality and momentum effects

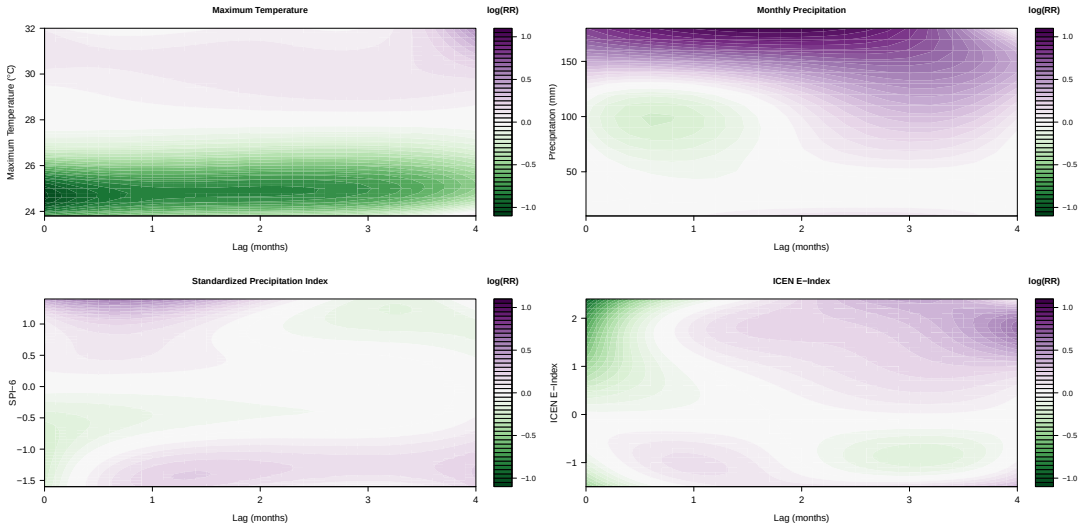


Figure SI B.2: **Omitting RSI and seasonality effects, exposure-lag-response relationships between climatic variables and dengue incidence.** Plots of relative risk (RR), on a logarithm scale, for the included DLNMs (Distributed Lag Nonlinear Models) in a climate-based Bayesian hierarchical model for dengue incidence fitted to the entire period of 140 months, where RR is defined relative to the risk induced by the mean observed value of each climate variable. Log RR values greater than 0 (pink to purple) correspond to heightened relative risk of dengue incidence, whilst values less than 0 (green) correspond to reduced RR. The four climatic variables were included in the model via DLNM specifications alongside spatial random effects, temporal random effects, and spatiotemporal random effects, but *not* including our usual fixed effects (of momentum and seasonality).

Climatic inference without maximum temperature effects

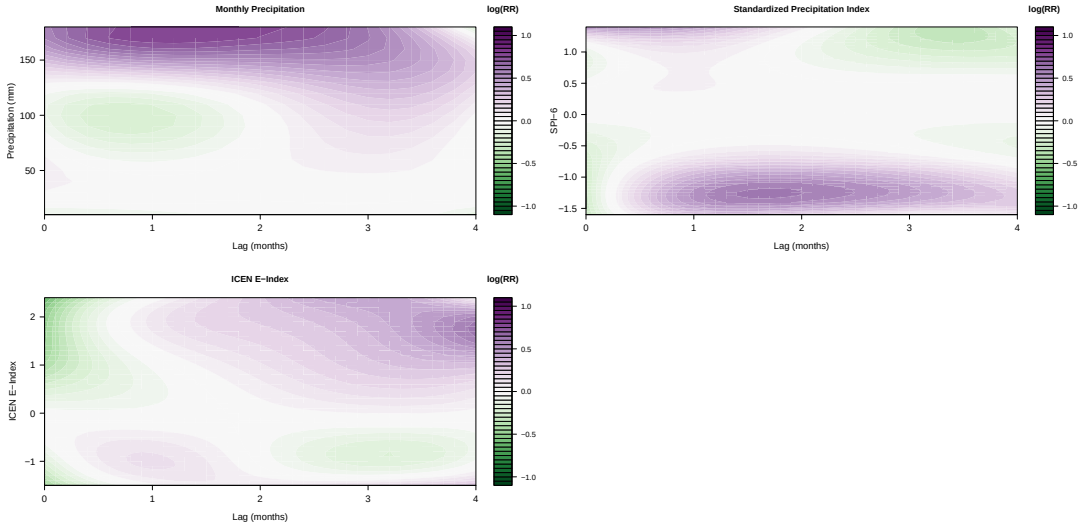


Figure SI B.3: **Omitting maximum temperature effects, exposure-lag-response relationships between climatic variables and dengue incidence.** Plots of relative risk (RR), on a logarithm scale, for the included DLNMs (Distributed Lag Nonlinear Models) in a climate-based Bayesian hierarchical model for dengue incidence fitted to the entire period of 140 months, where RR is defined relative to the risk induced by the mean observed value of each climate variable. Log RR values greater than 0 (pink to purple) correspond to heightened relative risk of dengue incidence, whilst values less than 0 (green) correspond to reduced RR. The three climatic variables were included in the model via DLNM specifications alongside spatial random effects, temporal random effects, and spatiotemporal random effects, and fixed effects (of momentum and seasonality).

Climatic inference without precipitation effects

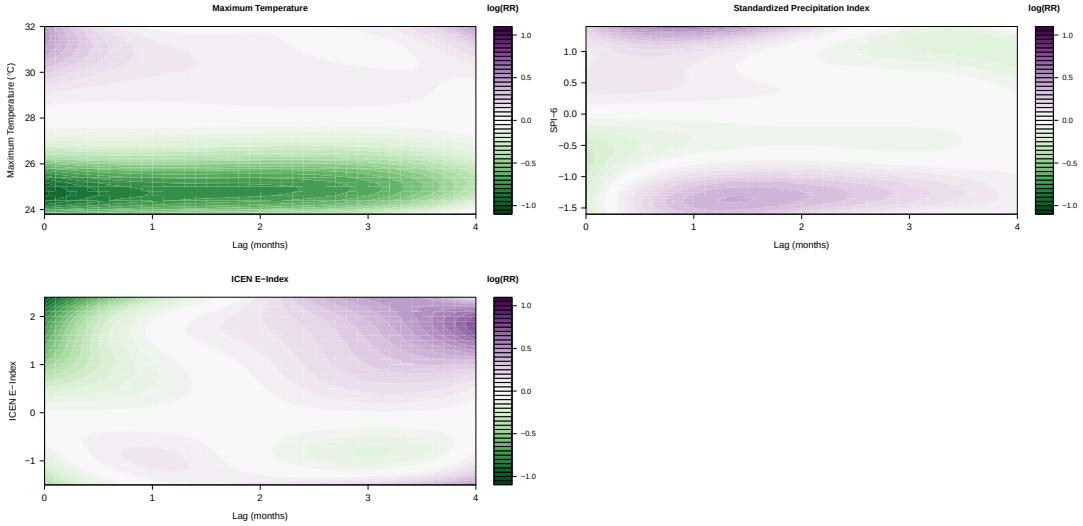


Figure SI B.4: **Omitting precipitation effects, exposure-lag-response relationships between climatic variables and dengue incidence.** Plots of relative risk (RR), on a logarithm scale, for the included DLNMs (Distributed Lag Nonlinear Models) in a climate-based Bayesian hierarchical model for dengue incidence fitted to the entire period of 140 months, where RR is defined relative to the risk induced by the mean observed value of each climate variable. Log RR values greater than 0 (pink to purple) correspond to heightened relative risk of dengue incidence, whilst values less than 0 (green) correspond to reduced RR. The three climatic variables were included in the model via DLNM specifications alongside spatial random effects, temporal random effects, and spatiotemporal random effects, and fixed effects (of momentum and seasonality).

Climatic inference without drought indicator effects

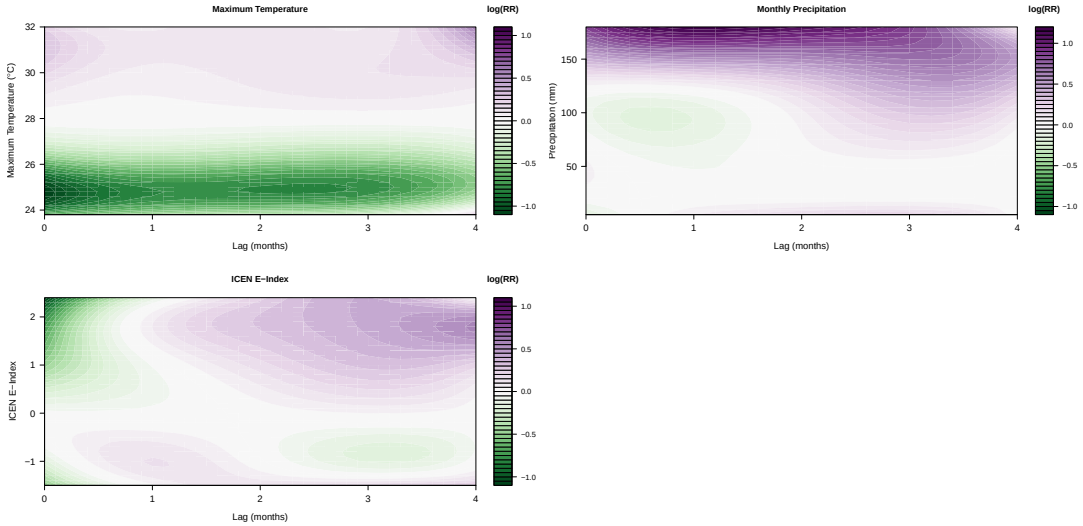


Figure SI B.5: **Omitting the drought indicator, exposure-lag-response relationships between climatic variables and dengue incidence.** Plots of relative risk (RR), on a logarithm scale, for the included DLNMs (Distributed Lag Nonlinear Models) in a climate-based Bayesian hierarchical model for dengue incidence fitted to the entire period of 140 months, where RR is defined relative to the risk induced by the mean observed value of each climate variable. Log RR values greater than 0 (pink to purple) correspond to heightened relative risk of dengue incidence, whilst values less than 0 (green) correspond to reduced RR. The three climatic variables were included in the model via DLNM specifications alongside spatial random effects, temporal random effects, and spatiotemporal random effects, and fixed effects (of momentum and seasonality).

Climatic inference without El Niño effects

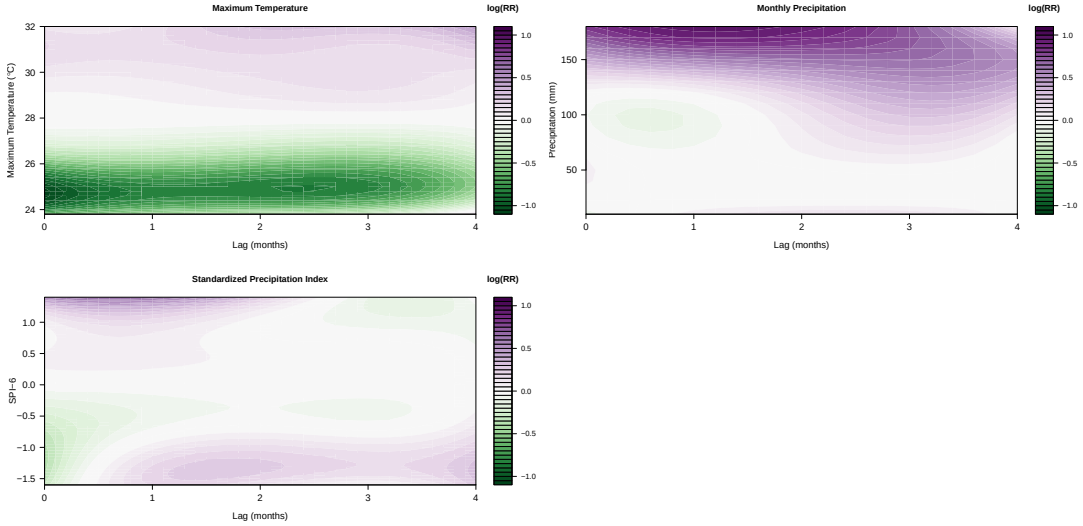


Figure SI B.6: **Omitting El Niño effects, exposure-lag-response relationships between climatic variables and dengue incidence.** Plots of relative risk (RR), on a logarithm scale, for the included DLNMs (Distributed Lag Nonlinear Models) in a climate-based Bayesian hierarchical model for dengue incidence fitted to the entire period of 140 months, where RR is defined relative to the risk induced by the mean observed value of each climate variable. Log RR values greater than 0 (pink to purple) correspond to heightened relative risk of dengue incidence, whilst values less than 0 (green) correspond to reduced RR. The three climatic variables were included in the model via DLNM specifications alongside spatial random effects, temporal random effects, and spatiotemporal random effects, and fixed effects (of momentum and seasonality).

B.4.3 Spatiotemporal climate trends

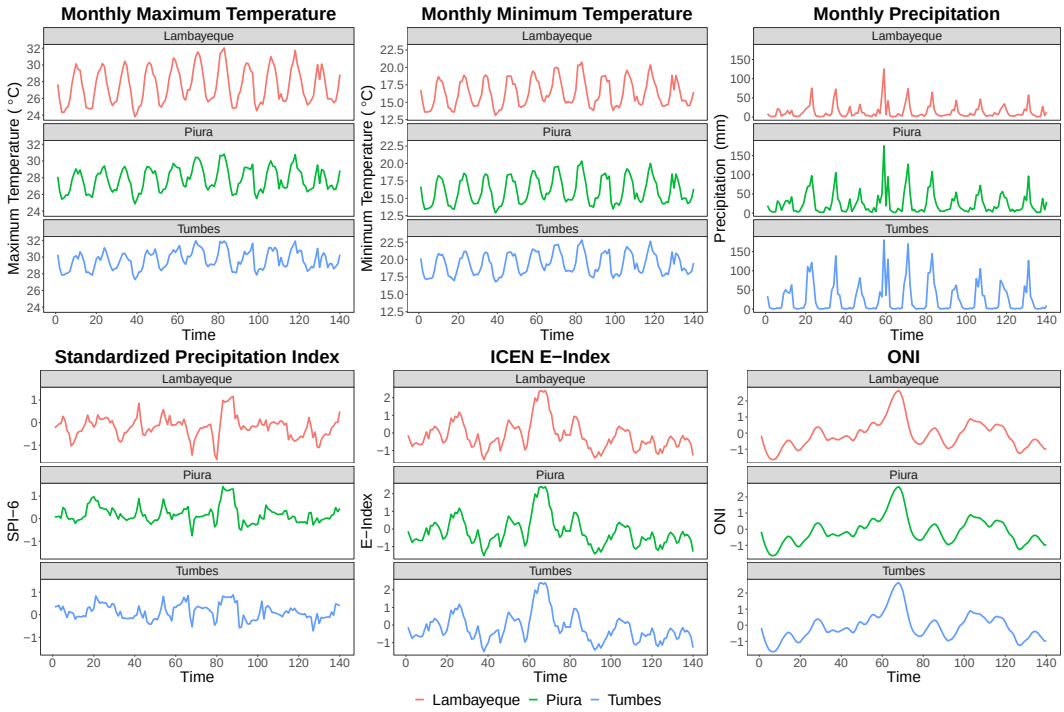


Figure SI B.7: **Historical climatic time series.** Exploratory plots which depict the monthly climatic data in Piura, Tumbes and Lambayeque over 140 months from May 2010 to December 2021. Both indicators for El Niño events (ICEN and ONI) are common across the three departments.

B.4.4 Seasonality in dengue incidence rates

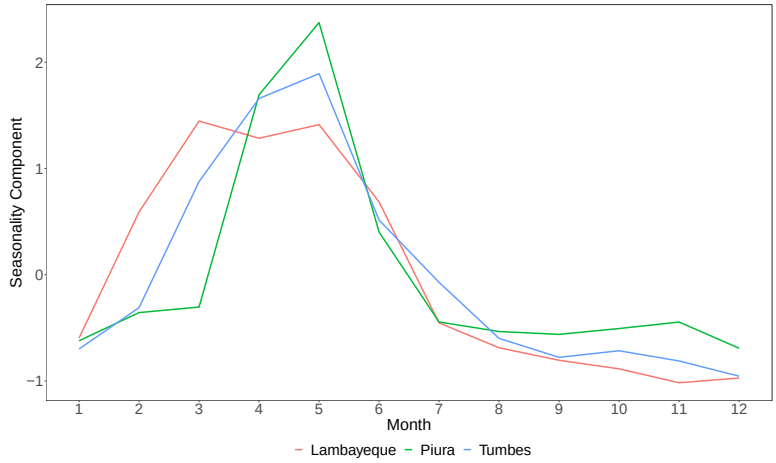


Figure SI B.8: **Department-specific seasonality.** A decomposition of the seasonality component in dengue incidence rates in each of the departments, calculated using the `bfast` package in R, which iteratively decomposes a time series into trend and seasonality components whilst detecting abrupt changes within the components.[516,517] The results of the `bfast` algorithm indicate similarity in the seasonal trends in dengue incidence rates across each of the three departments.

B.4.5 Cumulative risk induced by climatic variables

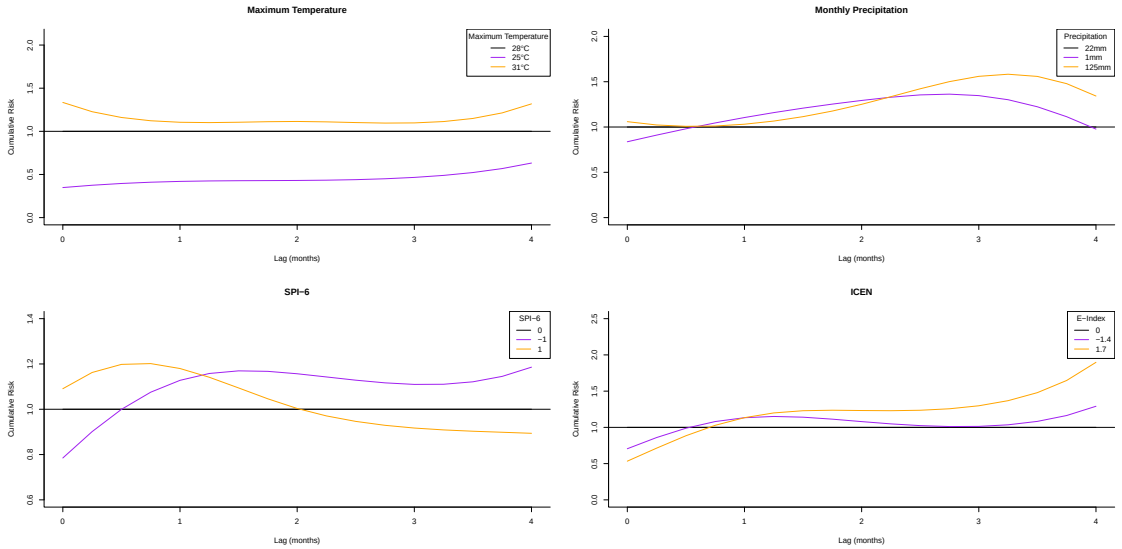


Figure SI B.9: **Cumulative association between climatic variables and DIR.** Plots of cumulative risk for the included DLNMs in climate-based Bayesian hierarchical model for dengue incidence, where cumulative risk is defined relative to the cumulative risk induced by the mean observed value of each climate variable. Cumulative risk values greater than 1 correspond to heightened cumulative risk of dengue incidence, whilst values less than 1 correspond to reduced cumulative risk. For example, as cumulative risk was negative for a monthly precipitation of 1mm at the current lag, then exposure over one month to dry conditions was associated with reduced risk, whilst positive values of cumulative risk at two months indicates that the exposure over two months resulted in a increased cumulative risk in dengue incidence. Note that in each sub-plot, we define a high and low value of the climatic variable, and the cumulative risk at each lag is defined relative to the cumulative risk induced by the mean value of the climatic variable.

B.4.6 Retrospective modelling in-sample predictive fit

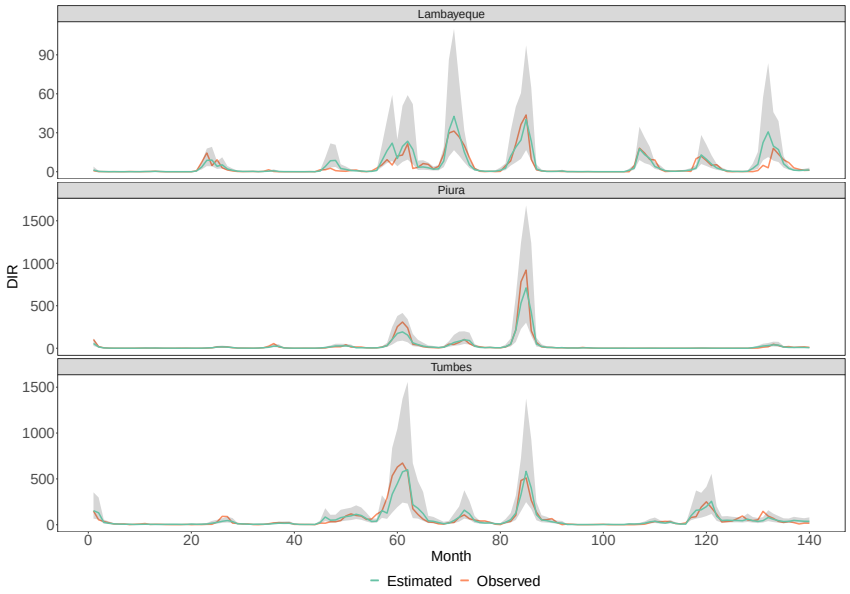


Figure SI B.10: **In-sample predictive performance of our Bayesian spatiotemporal model**. The DIR (Dengue Incidence Rate) time series (gold) for each department is plotted alongside the posterior median estimate (forest green) for each observation and the corresponding estimated 95% credible intervals (shaded grey). Figure 4.4 is the corresponding figure for the cross-validation setting which display the estimated out-of-sample predictive performance and thus, approximates the model's generalisability.

B.4.7 Estimated yearly random effects by department

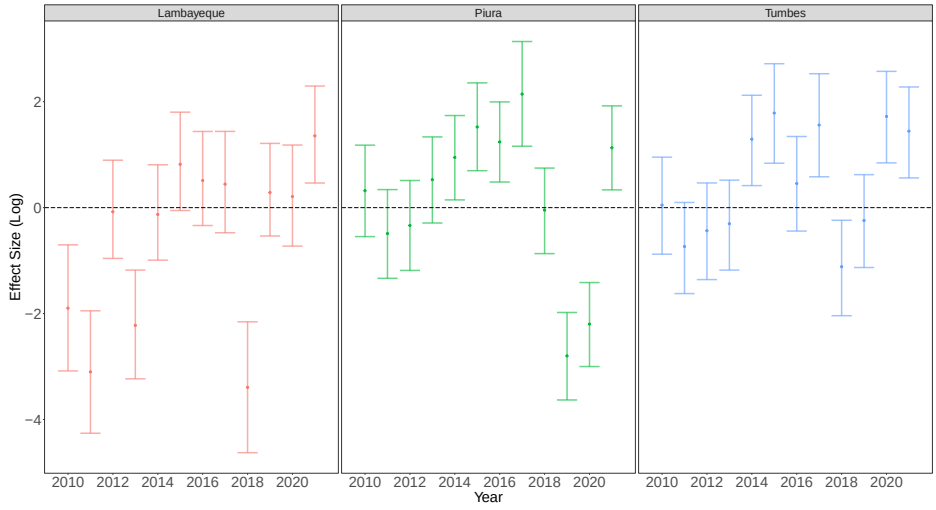


Figure SI B.11: **Estimated posterior distributions of department-specific yearly random effects**. The visualisation displays the posterior median estimate (circles) of the effect size (on a logarithmic scale) of the yearly random effect for each department, whilst the errorbars capture the 95% credible intervals. Note that the effect size can be interpreted as the contribution of the department-specific yearly random effect to the logarithm of DIR (Dengue Incidence Rate) observations within each year. The variability across departments in the yearly random effects is indicative of the differences in trends of year-to-year heterogeneities which prevented a shared yearly random effect (for each department) providing sufficient explanatory power.

B.4.8 Estimated monthly random effects by department

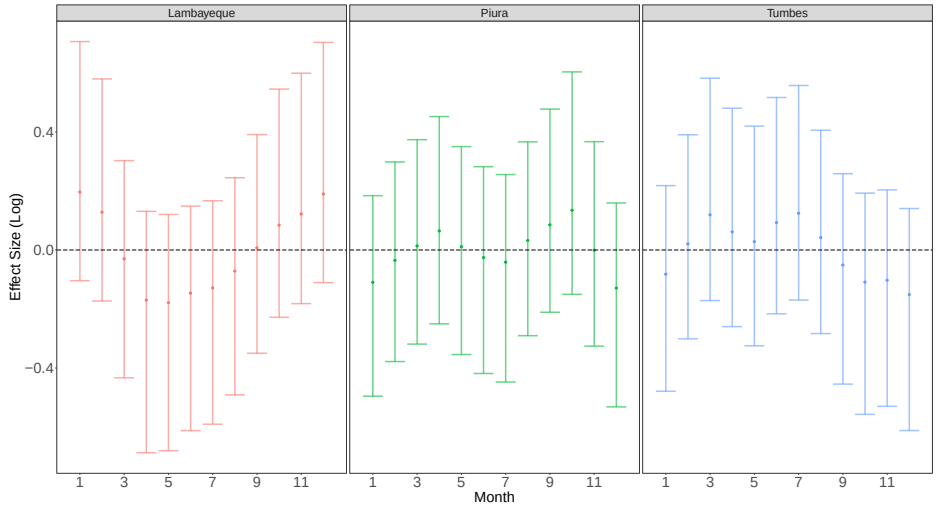


Figure SI B.12: **Estimated posterior distributions of department-specific monthly random effects**. The visualisation displays the posterior median estimate (circles) of the effect size (on a logarithmic scale) of the monthly random effect for each department, whilst the errorbars capture the 95% credible intervals. Note that the effect size can be interpreted as the contribution of the department-specific monthly random effect to the logarithm of DIR (Dengue Incidence Rate) observations across the years. Note that the contribution to DIR is in addition to the common seasonality indicator variable for summer months.

B.4.9 Estimated spatially structured random effects by department

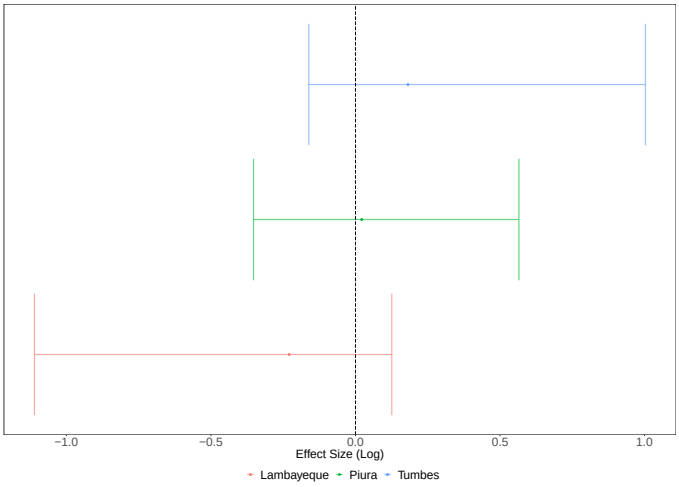


Figure SI B.13: **Estimated posterior distributions of department-specific structured spatial component effects**. The visualisation displays the posterior median estimate (circles) of the effect size (on a logarithmic scale) of the spatially structured random effect component of the BYM2 model for each department, whilst the errorbars capture the corresponding 95% credible intervals. Note that the effect size can be interpreted as the contribution of the department-specific spatial random effect to the logarithm of DIR (Dengue Incidence Rate) observations across the years.

B.4.10 Estimated effect of momentum indicator

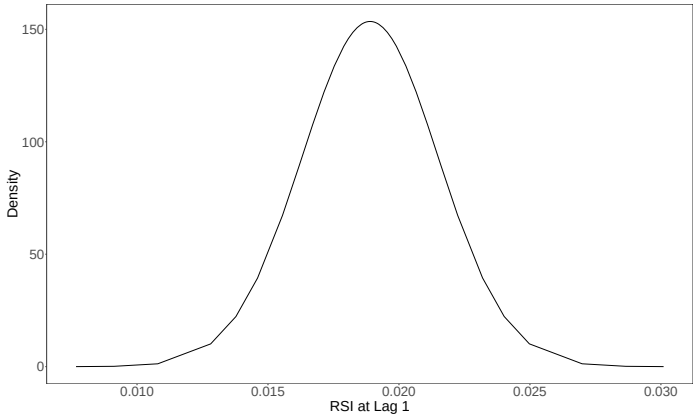


Figure SI B.14: **Estimated posterior distributions effects of momentum indicator**. The visualisation displays the posterior distribution of the effect size (on a logarithmic scale) of the lagged momentum indicator. Note that the effect size can be interpreted as the contribution of the lagged momentum indicator, calculated for three months of DIR (Dengue Incidence Rate) observations, to the current logarithm of DIR. We use the lagged momentum indicator as this ensures that the covariate does not include the response (DIR) itself.

B.4.11 Visualisation of climate-based out-of-sample errors

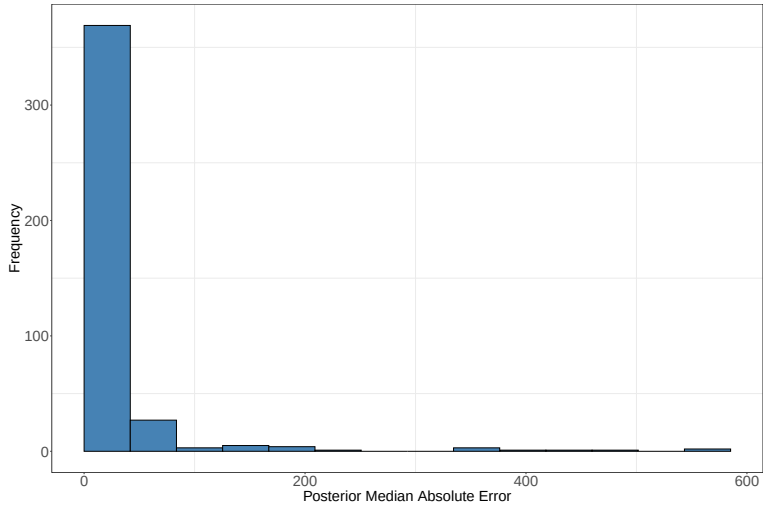


Figure SI B.15: **Leave-one-out cross-validation errors visualisation**. The histogram depicts the posterior median estimate of the absolute error for each observation, which was derived individually for observations by computing the median absolute difference between the posterior predictive value and the corresponding observed DIR (Dengue Incidence Rate). 89.6% of the posterior median estimates (n = 180) for the absolute forecast errors are less than 50 per 100,000.

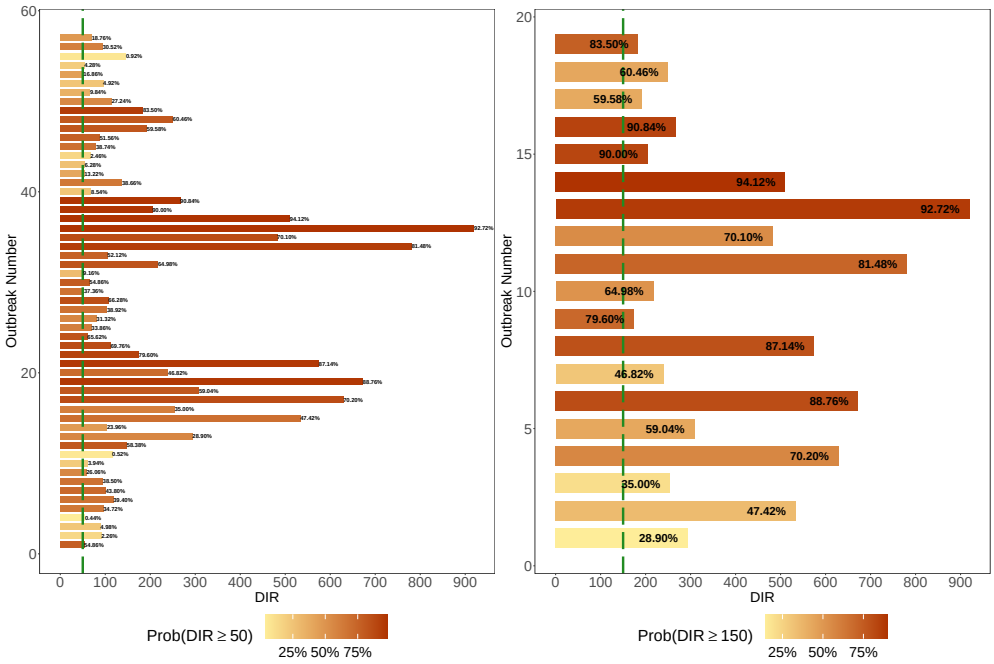


Figure SI B.16: **Retrospective modelling outbreak detection.** Among the observations with DIR (Dengue Incidence Rate) greater than 50 (left) and greater than 150 (right) per 100,000, the plots depict posterior probabilities of DIR exceeding thresholds (green) of 50 and 150 respectively, obtained via a leave-one-time-point-out fitting of models. The plots capture the model framework's sensitivity in estimating substantial out-of-sample dengue outbreaks. Within the two subsets of observations, the median posterior probabilities of exceeding 50 per 100,000 and 150 per 100,000 are 0·68 and 0·57 respectively. To also display the false positive rate and thus ensure a representative picture of our model's outbreak detection capabilities, Figure SI B.17 is the analogous visualisation for the corresponding posterior probabilities of the observations with DIR less than the thresholds of 50 and 150 per 100,000.

Table SI B.3: **Summary statistics for LOOCV outbreak detection performance.** Out-of-sample observations (n = 417), across 2010 to 2021, were classified as being predicted true outbreaks such that DIR exceeded 50 (or 150) per 100,000 if the posterior probability of DIR exceeding 50 (or 150) per 100,000 was greater than a calibrated cut-off probability of 0.16 (or 0.13). *True Positive* measures the hit rate, or equivalently the proportion of true outbreaks correctly detected, whilst *False Positive* measures the proportion of non-outbreaks incorrectly classified as outbreaks. *Accuracy* represents the proportion of observations whose outbreak classification matched the true observed state (such as a predicted outbreak coinciding with an outbreak). *Precision* measures the proportion of predicted outbreaks which were true outbreaks. Finally, the *AUC* is the area under the Receiver Operating Characteristic (ROC) curve, is used as a measure of our model's skill for distinguishing between outbreaks and non-outbreaks, and here, represents an upper bound on the model's retrospective performance (due to calibration of our cut-off probability for outbreak classification). 95% confidence intervals (CIs) for summary statistics were derived using exact binomial confidence limits [482,522], with the exception of the 95% CI for AUC being calculated using 2,000 bootstrap replicates.

Outbreak Threshold	True Positive	False Positive	Accuracy	Precision	AUC
DIR ≥ 50	0·93	0·13	0·88	0·54	0·90
95% CI	(0·83, 0·98)	(0·09, 0·16)	(0·85, 0·91)	(0·44, 0·64)	(0·87, 0·94)
DIR ≥ 150	1·00	0·06	0·94	0·43	0·97
95% CI	(0·82, 1·00)	(0·04, 0·09)	(0·91, 0·96)	(0·28, 0·59)	(0·96, 0·98)

B.4.12 Reliability of retrospective modelling outbreak detection

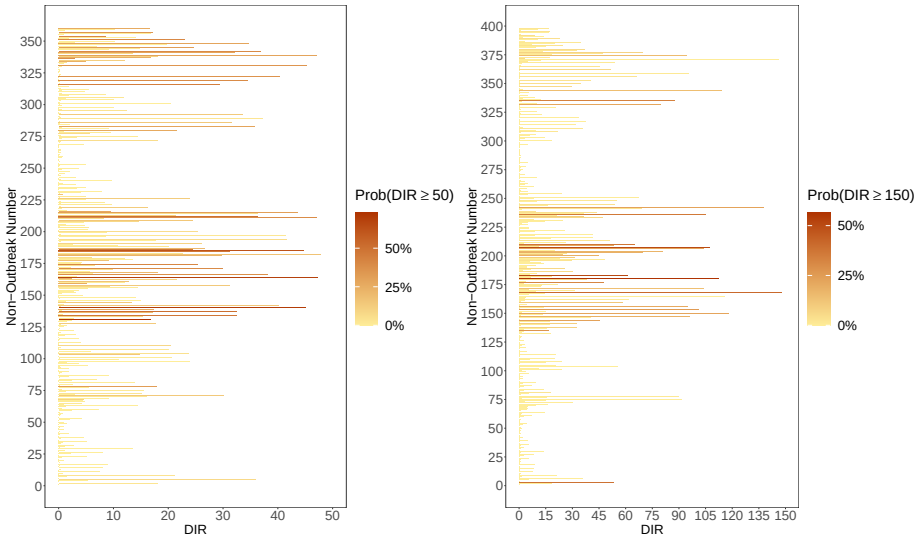


Figure SI B.17: **Reliability of estimated out-of-sample outbreaks.** Among the observations with DIR (Dengue Incidence Rate) less than 50 (left) and DIR less than 100 (right), the plots depict posterior probabilities of forecasted DIR exceeding thresholds of 50 per 100,000 (left) and 150 per 100,000 (right). The plots visualise the model's false positive rate in terms of forecasting severe dengue outbreaks. In the subset of observations with DIR less than 50, the mean posterior probability of exceeding 50 is 0·05, whilst the maximum posterior probability of exceeding 50 is 0·60. Similarly, in the subset of observations with DIR less than 150, the mean posterior probability of exceeding 50 is 0·08, whilst the maximum posterior probability of exceeding 50 is 0·43. In terms of the true positive rate (or hit rate), Figure SI B.16 is the analogous visualisation for the corresponding posterior probabilities of the observations with DIR greater than the thresholds of 50 and 150.

B.4.13 Visualisation of climate-based forecasting errors

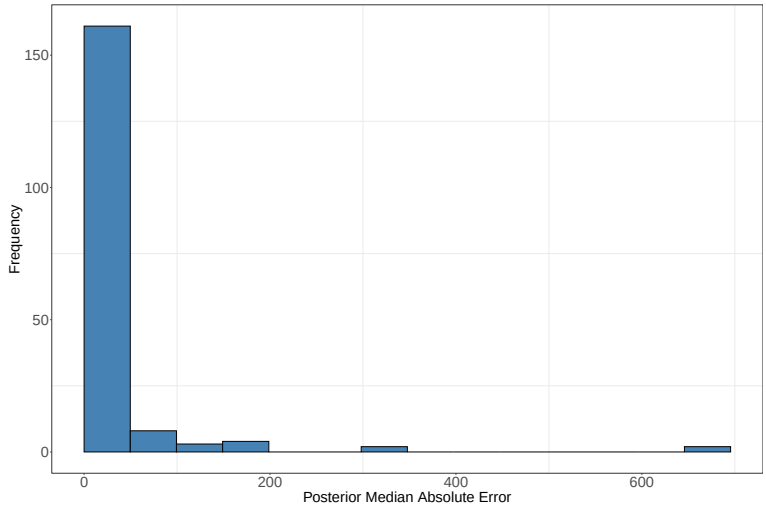


Figure SI B.18: **Forecasting errors visualisation**. The histogram depicts the posterior median estimate of the absolute error for each observation, which was derived individually for observations by computing the median absolute difference between the posterior predictive value and the corresponding observed DIR. 89.2% of the posterior median estimates (n = 180) for the absolute forecast errors are less than 50 per 100,000.

B.4.14 Reliability of forecasting outbreak detection

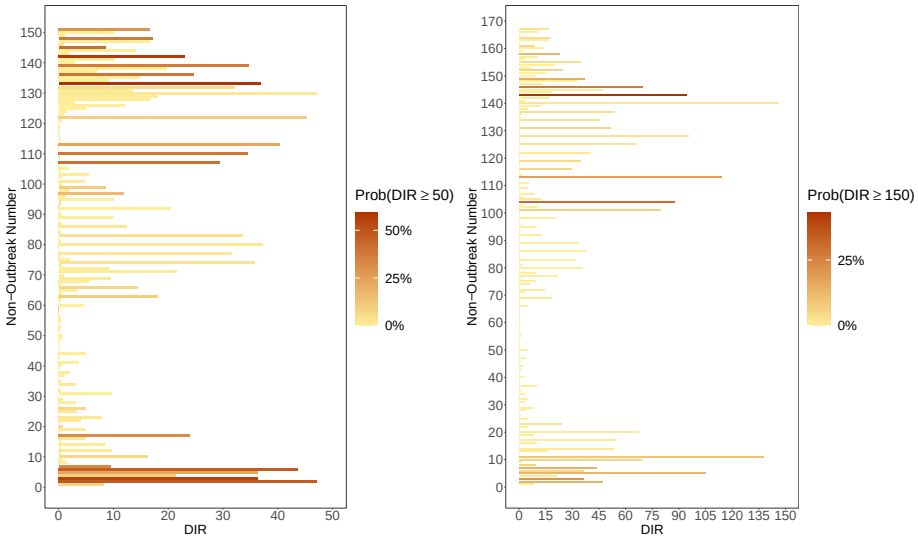


Figure SI B.19: **Reliability of forecasted outbreak detections.** Among the observations with DIR less than 50 (left) and DIR less than 100 (right), the plots depict posterior probabilities of forecasted DIR exceeding thresholds of 50 per 100,000 (left) and 150 per 100,000 (right). The plots visualise the model's false positive rate in terms of forecasting severe dengue outbreaks. In the subset of observations with DIR less than 50, the mean posterior probability of exceeding 50 is 0·05, whilst the maximum posterior probability of exceeding 50 is 0·59. Similarly, in the subset of observations with DIR less than 150, the mean posterior probability of exceeding 150 is 0·02, whilst the maximum posterior probability of exceeding 50 is 0·43. In terms of the true positive rate (or hit rate), Figure 4.6 is the analogous visualisation for the corresponding posterior probabilities of the observations with DIR greater than the thresholds of 50 and 150.

B.4.15 Non-model-based simple forecasting outbreak detection

Table SI B.4: **Summary statistics for non-model-based forecasting outbreak detection across 2018 to 2021.** DIR (Dengue Incidence Rate) observations (n = 180) were classified one month ahead of time as being predicted true outbreaks with DIR exceeding 50 (or 150) per 100,000 *only* if the previous month's DIR had exceeded 50 (or 150) per 100,000. The performance metrics below, and associated 95% confidence intervals (CIs), each have the same interpretation as defined in Table SI B.3. Classification of outbreaks below can be compared to our model-based classification of future outbreaks in Table 4.1, where 100% of all outbreaks with DIR ≥ 50 (or 150) per 100,000 were correctly forecasted one month ahead of time, alongside higher overall forecasting skill implied by larger Area Under the Curve (AUC) values.

Outbreak Threshold	True Positive	False Positive	Accuracy	Precision	AUC
DIR ≥ 50	0·80	0·03	0·94	0·80	0·89
(95% CI)	(0·61, 0·93)	(0·01, 0·07)	(0·90, 0·97)	(0·61, 0·93)	(0·81, 0·97)
DIR ≥ 150	0·70	0·02	0·97	0·70	0·84
95% CI	(0·35, 0·93)	(0·00, 0·05)	(0·91, 0·98)	(0·35, 0·93)	(0·69, 0·99)

Appendix C

Appendix — Multi-model approach to understand and predict past and future dengue epidemic dynamics

C.1 Acronyms

Table SI C.1: **Summary of acronyms used throughout the second manuscript of Chapter 4.**

Abbreviation	Definition
Ae. Aegypti	*Aedes Aegypti*
ACF	Autocorrelation Function
AUC	Area Under the Curve
BYM	Besag, York and Mollié
CRPS	Continuous Rank Probability Score
CV	Cross-Validated
DIC	Deviance Information Criterion
DIR	Dengue Incidence Rate
DLNM	Distributed Lag Nonlinear Models
ENSO	El Niño Southern Oscillation
ICEN	El Niño Coastal Index
IGP	Geophysical Institute of Peru
INEI	National Institute of Statistics and Information of Peru
INLA	Integrated Nested Laplace Approximation
IQR	Interquartile Range
ONI	Oceanic Niño Index
prec	Precipitation
Q1	Lower Quartile
Q3	Upper Quartile
ROC	Receiver Operating Characteristic
RMSE	Root-Mean Square Error
RR	Relative Risk
RSI	Relative Strength Index
RW	Random Walk
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
SPI-6	Standardized Precipitation Index (6-Month)
SST	Sea Surface Temperature
tmax	Maximum Temperature
tmin	Minimum Temperature
WIS	Weighted Interval Score
WHO	World Health Organization

C.2 Ensemble model abbreviations

C.2.1 Forecasting models

Table SI C.2: **Quantile-based ensemble framework abbreviations used in figures and tables.**

Model Abbreviation	Description
Baseline	Epidemiologically naive random walk model that uses the most recent month's observed cases as a forecast for the next month.
Bayes-Climate	Bayesian spatiotemporal model with nonlinear, delayed climatic effects, spatial random effects, a momentum indicator, and temporal effects
SARIMA	Seasonal Auto-Regressive Integrated Moving Average model specified for each province independently at each time
TCN	Temporal Convolutional Network with climatic covariates specified for each province independently
TimeGPT	Transformer-based foundational time series forecasting model with climatic covariates specified for each province independently
TimeGPT-NoCov	Transformer-based foundational time series forecasting model without any covariates specified for each province independently
EW-Mean *	Equally weighted mean average of all individual forecasting models
EW-Mean-NoBase *	Equally weighted mean average of all individual forecasting models, excluding the baseline model
EW-Mean-NoBayes *	Equally weighted mean average of all individual forecasting models, excluding the Bayesian climate-based model
EW-Mean-NoCov *	Equally weighted mean average of all covariate-free individual forecasting models. This excludes covariate-based models – the Bayesian climate-based, TimeGPT, and TCN models.
Median *	Median of all individual forecasting models
Median-NoBase	Median of all individual forecasting models, excluding the baseline model
Median-NoBayes *	Median of individual forecasting models, excluding the Bayesian climate-based model
Median-NoCov *	Median of individual forecasting models, excluding covariate-based models (the Bayesian climate-based, TimeGPT, and TCN models)
Prov-Trained*	Trained ensemble with province-dependent weights determined to minimise the sum of the quantile losses
Trained*	Trained ensemble with spatially homogeneous weights determined to minimise the sum of the quantile losses

C.3 EPIFORGE 2020 guidelines

Table SI C.3: **Checklist for EPIFORGE 2020 guidelines:** We provide details of our reporting for the checklist items of EPIFORGE 2020[564]. Page numbers pertain to original manuscript[46].

Section of Manuscript	#	Checklist Item	Reported on Page
Title/Abstract	1	Describe the study as forecast or prediction research in at least the title or abstract.	1
Introduction	2	Define the purpose of study and forecasting targets.	1-2
Methods	3	Fully document the methods.	3-7
Methods	4	Identify whether the forecast was performed prospectively, in real time, and/or retrospectively.	5
Methods	5	Explicitly describe the origin of input source data, with references.	3
Methods	6	Provide source data with publication, or document reasons as to why this was not possible.	17
Methods	7	Describe input data processing procedures in detail.	3
Methods	8	State and describe the model type, and document model assumptions, including references.	4-7
Methods	9	Make the model code available or document the reasons why this is not possible.	17
Methods	10	Describe the model validation and justify the approach.	4-7
Methods	11	Describe the forecast accuracy evaluation method used, with justification.	5-6
Methods	12	Where possible, compare results to a benchmark or other comparator model, with justification of comparator choice.	6
Methods	13	Describe the forecast horizon, with justification of its length.	5
Results	14	Present and explain uncertainty of forecasting results.	10-11
Results	15	Briefly summarize the results in nontechnical terms, including a non-technical interpretation of forecast uncertainty.	10-11
Results	16	If results are published as a data object, encourage a time-stamped version number.	n/a
Discussion	17	Describe the weaknesses of the forecast, including weaknesses specific to data quality and methods.	15-16
Discussion	18	If the forecast research is applicable to a specific epidemic, comment on its potential implications and impact for public health action and decision-making.	16-17
Discussion	19	If the forecast research is applicable to a specific epidemic, comment on how generalizable it may be across populations.	16-17

n/a= Not Applicable

C.4 Climate-based model: Additional details

The following section provides additional details to the description of the modelling framework presented in Section 4.2.2.

Model formula

Let y_{it} be the reported province-level ($i = 1, \dots 14$) monthly ($t = 1, \dots 140$)) case counts of dengue from 2010 to 2021. Then, y_{it} was modelled with a Zero-Inflated (ZI) Poisson distribution to account for the likely presence of excess zero counts in the province-level spatial resolution.

$$\begin{aligned} y_{it} &\sim ZIPoisson(\lambda_{it}, \pi) \\ \log(\lambda_{it}) &= \log(P_{it}) + \log(\eta_{it}) \\ \log(\eta_{it}) &= \beta_0 + \boldsymbol{x_i\beta} + \gamma_{i,m(t)} + \delta_{i,a(t)} + u_i + v_i + \sum_{l=1}^{L} bs(c_{lit}, n_k^{(l)}, 4) \end{aligned} \tag{C.4.1}$$

where λ_{it} denotes the rate parameter of the Poisson distribution and π is the probability of a zero count. Similar to our previous work[45], we included a population offset $\log(P_{it})$ such that $\log(\eta_{it})$ represented the DIR per 100,000 population. The DIR was then modelled with an intercept β_0, alongside $L = 4$ delayed climatic effects ($bs(c_{lit})$) with $n_k^{(l)}$ knots for the k^{th} climatic variable, fixed effects, temporal random effects ($\gamma_{i,m(t)}$ and $\delta_{i,a(t)}$ for monthly and yearly effects respectively), and spatial random effects (u_i and v_i). We will now describe these fixed and random effects in further detail.

C.4.1 Delayed nonlinear climatic effects

We used distributed lag nonlinear models (DLNMs) to capture the potentially nonlinear and delayed effects of climatic variables on dengue incidence rates. As described in[45], DLNMs are a flexible modelling framework which are based on the statistical concept of a *crossbasis*; a bi-dimensional functional space comprised of a combination of two sets of basis functions which represent the relationships with the response in the dimensions of the predictor and lags respectively[394]. In particular, usage of the DLNM framework involves the simultaneous specification of i) nonlinearity in the exposure-response relationship and ii) delayed effects in the lag-response dimension. The specification is achieved via the *cross-basis*, thus creating an exposure-lag-response function.

We used the R package `dlnms` to estimate our DLNMs. Here, similar to[45], we employed basis splines (B-splines) for our two sets of basis functions, which were combined to generate our cross-basis functions[394]. For each climatic variable, in the exposure-response dimension, we specified two equally spaced knots to allow for flexibility (and hence, complexity) in the relationship between the climatic variable and dengue incidence. The basis functions were centred on the mean average of the climatic variable. Then, we considered varying maximum time lags (of between two to four months) for the lag-response dimension for each climatic variable, and specified the knots at individual months to reflect the exploratory analyses.

C.4.2 Temporal and spatial random Effects

The current model's temporal and spatial random effects have been described elsewhere in our previous analysis[45], except here we replaced the department-specific temporal effects with corresponding province-specific (monthly and yearly) temporal effects and the department-level BYM2 model (an adapted Bayesian version of the Besag-York-Mollié (BYM) model) with a province-level BYM2 model[519]. Similar to previous analyses which implemented the BYM2 model, penalised complexity prior distributions were employed for precision parameters (τ), which meant that $P(\frac{1}{\tau} > 0 \cdot 5) = 0 \cdot 01$[78,520].

Briefly, the monthly and yearly random effects aimed to capture seasonal (cyclic) patterns in a province's DIRs and year-to-year heterogeneities (such as time-varying reporting rates). The BYM2 model aimed to capture the spatial autocorrelation between the individual provinces, whilst additionally estimated any unstructured (i.e. non-spatial) heterogeneity in the provinces.

C.4.3 Fixed Effects

Our fixed effects, expressed as a matrix of four columns $\boldsymbol{x_i}$ in Equation C.4.1, comprised of; i) our seasonality indicator for the summer months (= 1 for months of December to April, and 0 otherwise) common to each department and ii) our momentum oscillator; Relative Strength Index (RSI), (which combined lagged RSI values of the past two months), iii) the proportion of a province's population declaring to be living in an urban area, and iv) the preceding month's year-to-year difference in DIR.

First, the objective of the simple binary seasonality variable was to capture seasonal trends of the summer months which are common across the departments. Whilst our climatic variables accounted for the influence of environmental conditions on vector abundance and incidence rates, the seasonality variable provided a means of accounting for possible population transience and increased outdoors activity during summer months.

Second, the proposed concept of the RSI variable was motivated by its usage as a technical indicator in financial markets to measure the strength and momentum of price movements[518]. For each department, we employed the past two values of the RSI, which were each calculated the smoothed moving average of the previous three month's DIRs. The formula used for RSI was:

$$\mathrm{RSI}_t = 100 - -\frac{100}{(1+\mathrm{RS})}$$

where RS is the *relative strength*, defined as the smoothed ratio of average gains (sum of the upward DIR movements over the past three months, divided by three – the number of months) over the average losses (sum of the downward DIR movements over the past three months, divided by three). By using the product of the past two RSI values ($\mathrm{RSI}_t \times \mathrm{RSI}_{t-1}$), we tracked i) the recent momentum in large DIR movements and ii) the overall magnitude (or recent levels) of DIR (via the product). So, even if we have recently observed downward movements in DIR, two strong consecutive momentum values would indicate that the DIR has recently been (or is currently) at high levels.

Third, the continuous urban variable was included to estimate the possible amplifying effect of a highly urbanised population on DIR, possibly arising due to greater human-to-vector contact and greater opportunities for larval habitat creation. This was a fixed variable over calendar time, indicating the percentage of the population living in urban (as opposed to rural) area.

Finally, we included the preceding month's year-to-year difference in DIR as a proxy for possible outbreak years where larger proportions of the population may have entered the susceptible sub-population. We modelled the effect using a natural spline with four knots to allow for sufficient flexibility for the likely nonlinearity in both the negative and positive effects in non-outbreak and outbreak years respectively. For example, if the preceding month's DIR was far (slightly) greater than the preceding year, we would expect the current month's DIR would be far (slightly) greater than the previous year.

C.4.4 Covariates

Table SI C.4: **Covariates used in our Bayesian climate-based model**. All covariates were determined by exploratory analyses and performance of the Bayesian climate-based model in the model development period of 2010-2017.

Covariate	Description
Monthly random effect	Cyclic random walk (of order 1) prior distribution used for monthly random effects for each province.
Yearly random effect	Exchangeable Normal(0,1) prior distribution used for yearly random effects for each province.
BYM2 spatial effects	Structured spatial and unstructured components to capture spatial correlation and individual province uniqueness respectively in an adapted Bayesian version of the Besag-York-Mollié (BYM) model with penalised complexity prior distributions [519,520].
Minimum temperature (*tmin*)	Two-month rolling average (right-aligned) of the monthly average of daily minimum temperature, specified using a DLNM (Distributed Lag Nonlinear Model). The DLNM featured B-splines with two equally-spaced knots in the exposure-response dimension and knots at one and two months in the lag-response dimension (which included a maximum lag of three months).
Precipitation (*prec*)	Two-month rolling average (right-aligned) of the monthly total precipitation (mm), specified using a DLNM. The DLNM featured B-splines with two equally-spaced knots in the exposure-response dimension and knots at one and two months in the lag-response dimension (which included a maximum lag of three months).
Standardized Precipitation Index (SPI-6)	Drought indicator (SPI), measuring precipitation anomalies over a six month period [512], specified using a DLNM. The DLNM featured B-splines with two equally-spaced knots in the exposure-response dimension and knots at one and two months in the lag-response dimension (which included a maximum lag of two months).
El Niño Coastal Index (ICEN)	The ICEN, a sea-surface temperature indicator specific to Peru [513,514], specified using a DLNM. The DLNM featured B-splines with two equally-spaced knots in the exposure-response dimension and knots at one and two months in the lag-response dimension (which included a maximum lag of four months).
Seasonality	A binary summer-winter seasonality variable to account for the non-climatic effects of possible population transience and/or general DIR patterns (across the provinces) in the absence of other amplifying effects (such as climatic influences).
Urban population	The percentage of a population living in an urban (as opposed to rural) area, as reported by the 2017 national census. Note that we assumed a constant (as opposed to time-varying) value (from 2017) to allow us to estimate the overall effect of a highly urbanised province.
Lagged monthly difference with previous year	One-month lagged value of the difference between that month's DIR and the DIR from the corresponding month in the previous year. The effect was estimated using a natural spline with four knots to allow for the likely nonlinear effects of negative and positive differences, and the variable was included in an attempt to capture outbreak years (as opposed to non-outbreak years).
Relative Strength Index (RSI)	The product of monthly momentum indicator, RSI (calculated over a horizon of three months), lagged by one month and two months, which aimed to capture momentum movements. Whilst it may seem counter-intuitive to also square negative momentum, in our exploratory analysis, we found stronger correlation between DIR and the squared lag of RSI, which may be explained by large negative movements subsequent preceding higher DIR (possibly due to reporting artefacts or seasonal patterns).

C.5 Ensemble framework: Additional details

C.5.1 Background to probabilistic ensemble forecasting

A probabilistic ensemble forecasting model is a statistical machine learning approach which combines information from several forecasting models and assigns probabilities to individual events or targets. An ensemble approach is increasingly used in infectious disease forecasting due to frequent outperformance of forecasts from individual models, whilst often providing decision-makers with more consistent, stable information across both time and space[172,764]. In the context of forecasting dengue incidence, a study (in the cities of Iquitos, Peru and San Juan, Puerto Rico) found that ensemble approaches produced superior forecasting skill (as measured by proper scoring rules) over an eight-year period[169].

C.5.2 Trained and untrained ensemble frameworks

For quantile forecasts, we use the following equation to weight the components of each ensemble framework:

$$q_r^E = \sum_{k=1}^{K} w_k q_r^k \tag{C.5.1}$$

where w_k is defined as before and q_r^k is the r^{th} quantile value estimated by the k^{th} model component. We produced forecasts for 23 quantile levels (identical to those used in the COVID-19 Forecast Hub in the United States[433]). For trained weighting of the ensemble (E), weights w_k are determined such that they minimised the sum of quantile losses over the past 12 months. These weights were determined for all provinces jointly (i.e. province-independent model weights) and for each province independently (i.e. province-dependent weights). For untrained weighting of the ensemble, the weighted sum (in Equation C.5.1) is replaced with the median or mean average function applied again to individual modelling components' predictive quantiles. To evaluate probabilistic sharpness and bias, we used the weighted interval score (WIS), a proper scoring rule which is a discrete approximation to the CRPS[58,470]. We assessed WIS on a logarithmic scale to follow recommended procedures, and lower scores indicate better performance.

C.6 Ensemble component forecasters: Additional details

As described in the main text, we developed Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models. These statistical time series models were fitted independently to the logarithm of monthly cases for each province independently using the Auto-ARIMA implementation of the Statsforecasts package and Darts library in Python[561,562]. This implementation selects the order of the ARIMA model based on the AIC[464], and quantile forecasts are generated by assuming normality in forecast errors. We refitted the model at each individual time point to reflect real-world conditions as new information becomes available. Here, we highlight that the updating of our SARIMA model, including the order (i.e. its parameters), at each time step is an important task. This updating differs to past forecasting studies for dengue (e.g.,[169,555]) where the order of the ARIMA/SARIMA models were fixed across the testing set. We allow for this updating to i) ensure equitable comparison and ii) to allow the most recent data determine the order of the model that we should use. Indeed, for deep learning models (see below), one would always allow its weights (and potentially other hyperparameters) to be updated at each time point, and so we perform a similar procedure for the SARIMA model.

The first deep learning model that we implemented was TimeGPT, a transformer-based model with self-attention that has been pre-trained on over 100 billion data points from different domains. Briefly, the model uses an encoder-decoder structure with multiple layers, each consisting of residual connections (which allow bypassing certain layers) and layer normalisation (which helps to stabilise the model)[401,563]. We implemented this model for the logarithm of cases with

and without climatic covariates (two-month rolling averages of precipitation and minimum temperature)[401,563]. While the TimeGPT model allows for zero-shot inference, we used a fine-tuning process which refines the model's parameters to suit the task of forecasting dengue cases. This fine-tuning process consists of ten training iterations of model fitting to the training data to minimise forecasting error. Then, the updated TimeGPT model (after fine-tuning) readily produces quantile forecasts by using conformal inference — a distribution-free way to produce statistically rigorous prediction intervals[426].

Our second deep learning model was a temporal convolutional network (TCN) from the Darts library in Python[561]. Again, the TCN architecture consists of an encoder-decoder structure. The encoder has stacked dilated causal convolutional neural networks which aim to describe longer-term temporal dependencies in the data. The decoder uses a residual block and an output dense layer. For each observation, we used a lookback period (or input chunk length) of 48 months which means that we allowed for the past 48 months of data to be considered by the TCN model when making predictions. To avoid making potentially invalid distributional assumptions, we used quantile regression to directly estimate the quantiles of the predictive distribution. We again used climatic covariates of two-month rolling averages of precipitation and minimum temperature. These covariates were based on exploratory analyses from the model development period.

C.7 Outbreak detection: Additional details

C.7.1 Classification of DIR exceeding thresholds

Our modelling frameworks (see Materials and methods) each produced predictive distributions which we represented using predictive quantiles.

First, using the samples from the predictive distribution, we derived probabilities of DIR surpassing thresholds of 50 (and 150) per 100,000. Rather than using an arbitrary cut-off probability for classifying whether an outbreak would occur, we devised an optimisation scheme which maximised the historical performance of the forecasting model. Hence, when classifying whether DIR would surpass a specific threshold for each given month, a cut-off probability was calibrated using past (i.e. training) data that maximised the historical AUC, the area under the Receiver Operating Characteristic (ROC) curve. The procedure ensured that we used the cut-off probability that maximised the difference between the true positive rate and false alarm rate. The aim was to ensure that the classification would provide public health authorities with a more reliable forecasts of whether DIR would surpass the threshold. To reflect a real-world scenario, the cut-off probabilities were iteratively updated as new data became available.

C.8 Additional discussion of limitations

Another key limitation across the study is the quality of data available. For dengue surveillance data, we did not have access to data on public health interventions (e.g., concurrent vector control programmes) or reporting rates (e.g., number of participating healthcare centres and hospitals) which can help to gauge underreporting. However, years with greater DIRs (relative to historical averages) were often associated with lower average proportions of cases being confirmed (hence, more being classified as probable cases), which may indicate time-varying pressures on the healthcare centres and surveillance system, and provides further motivation for our modelling of year-to-year heterogeneities in our climate-based model (Figure SI C.42). The surveillance data also did not provide stratification by serotype, and the associated unavailability of seroprevalence data restricted us from accounting for potentially time-and-space-varying serotype dominance and immunity profiles of populations. Similarly, the lack of demographic data (such as age of case) for probable and confirmed cases limited us from assessing how mean age of infection influences the epidemic dynamics (such as the force of infection, as performed by[438]) or how the demographics relate to our computed wavelet-based measures (such as synchrony, coherence,

or the period of multiannual cycles). Thus, consistent with the WHO's cited lack of data on circulating serotypes[569], we advocate for future collation of age-structured dengue cases, alongside enhanced genomic and serological surveillance across at-risk areas, all of which would provide insights about immunology and virology issues (such as multitypic immunity, genomic diversity, and repeat infections) across space and time.

C.9 Figures and tables

C.9.1 Wavelet and climate-based analyses

Exploratory analyses

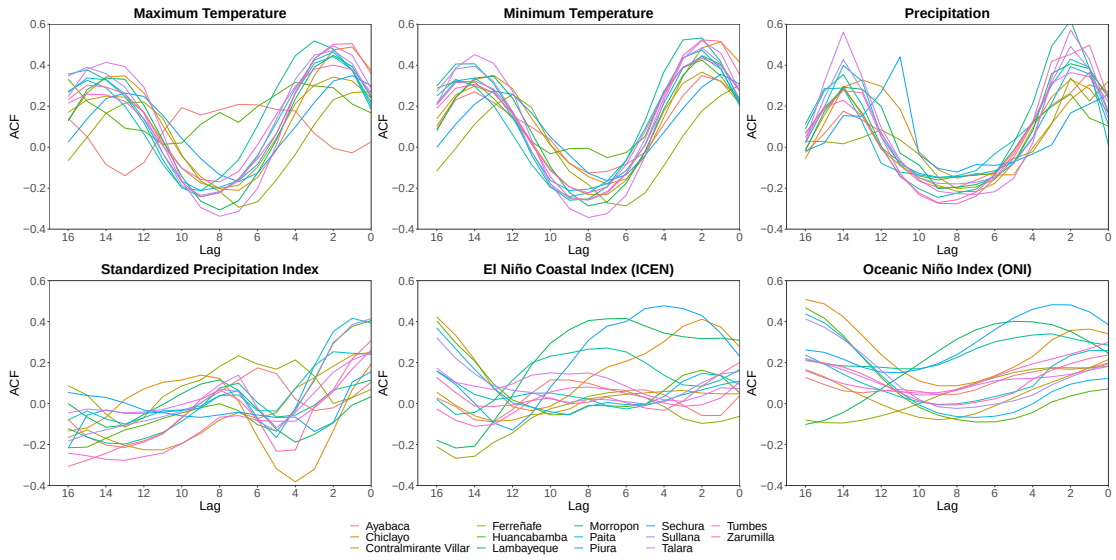


Figure SI C.1: **Relationships between lagged climatic variables and dengue incidence rates.** Across the model development period (2010 to 2017), the cross-correlation plots depict the province-level relationship between six individual climatic variables and the monthly dengue incidence rate (DIR) per 100,000. ACF (y-axis) denotes the auto-correlation function, and the lag (x-axis) denotes the lag (in months) from the climatic exposure to subsequent monthly DIR.

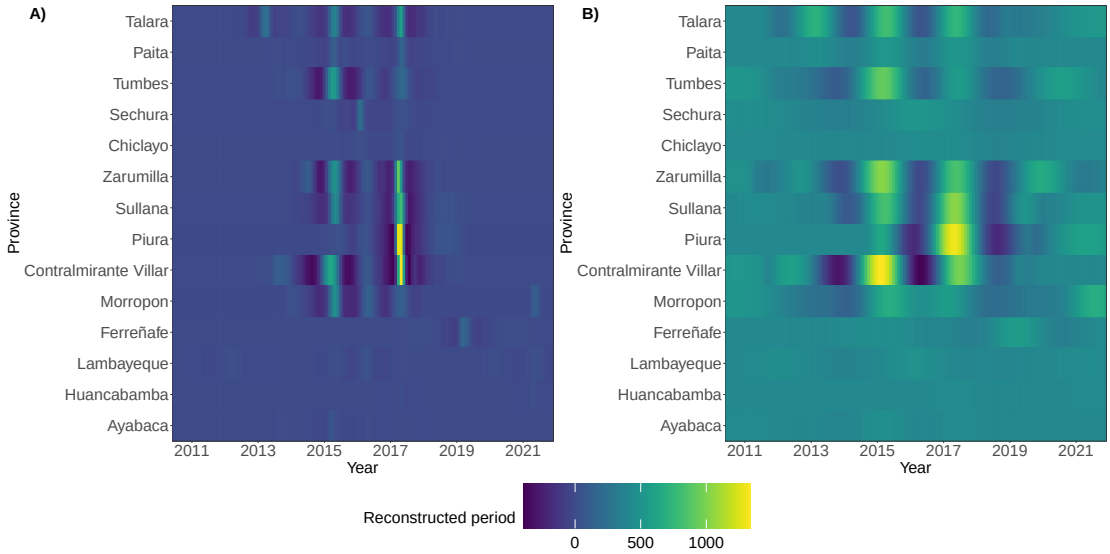


Figure SI C.2: **Wavelet reconstructions of cycles of Dengue Incidence Rate (DIR) time series.** Left: Reconstructed annual cycles of province-level DIRs, where provinces are sorted in ascending order by percentage of population living in urban areas. Right: Reconstructed multiannual cycles of province-level DIRs, where provinces are again sorted in ascending order by percentage of population living in urban areas. Higher values of the annual (multiannual) period indicate peaks in DIR across the annual (multiannual) reconstructions, where the wavelet-based reconstructions are the reconstructions of province-level DIR time series using the initial decompositions into time-frequency and then, frequencies corresponding to annual (multiannual) cycles.

Spatiotemporal incidence trends and properties of reconstructed cycles

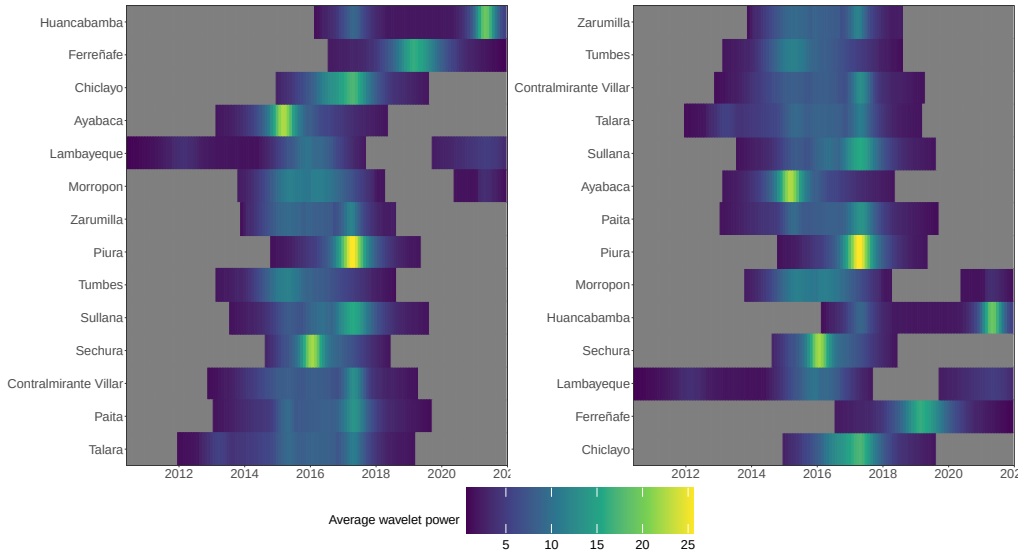


Figure SI C.3: **Average wavelet power of dengue incidence.** Cells depict average wavelet power of annual cycles per time step in provinces **(A)** sorted by longitude and **(B)** sorted by latitude, where non-filled (grey) cells depict non-significant wavelet power.

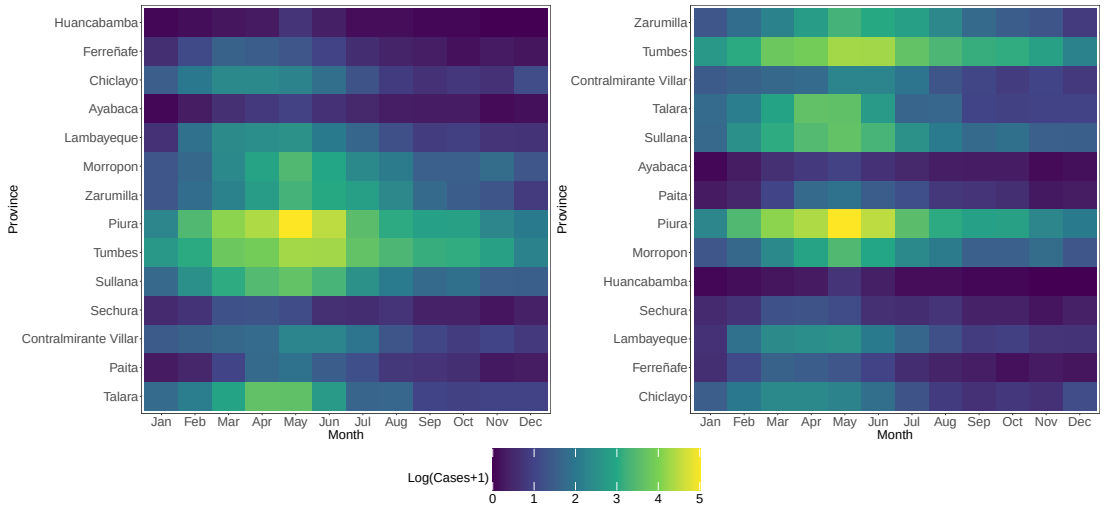


Figure SI C.4: **Average spatiotemporal trends in dengue epidemics.** Monthly mean average of provinces' log(Cases + 1) where provinces are sorted **(A)** by longitude (from west to east) and **(B)** by latitude (from south to north). We visualise the logarithm of cases to enable comparison of epidemics of different sizes across different geographies. This is aligned with our proper scoring in our forecasting, as recommended by recent guidelines[213].

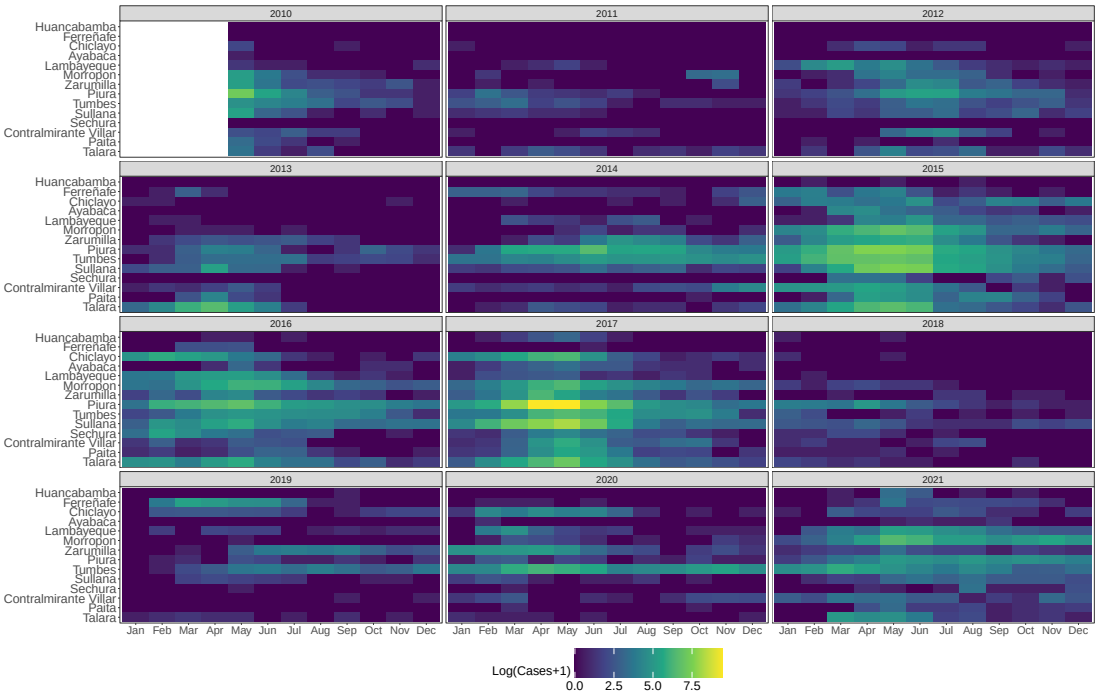


Figure SI C.5: **Longitudinal spatiotemporal trends in dengue epidemics.** Provinces' log(Cases + 1) are shown, where provinces are sorted by longitude (from west to east). We visualise the logarithm of cases to enable comparison of epidemics of different sizes across different geographies. This is aligned with our proper scoring in our forecasting, as recommended by recent guidelines[213].

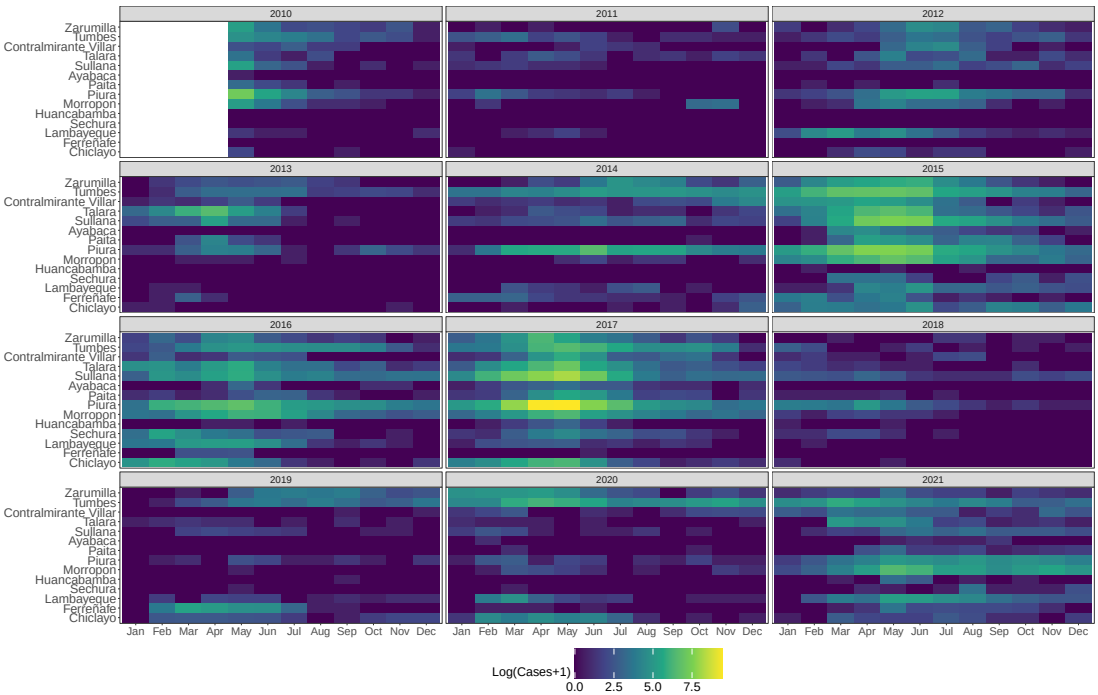


Figure SI C.6: **Latitudinal spatiotemporal trends in dengue epidemics.** Provinces' log(Cases + 1) are shown, where provinces are sorted by latitude (from south to north). We visualise the logarithm of cases to enable comparison of epidemics of different sizes across different geographies. This is aligned with our proper scoring in our forecasting, as recommended by recent guidelines[213].

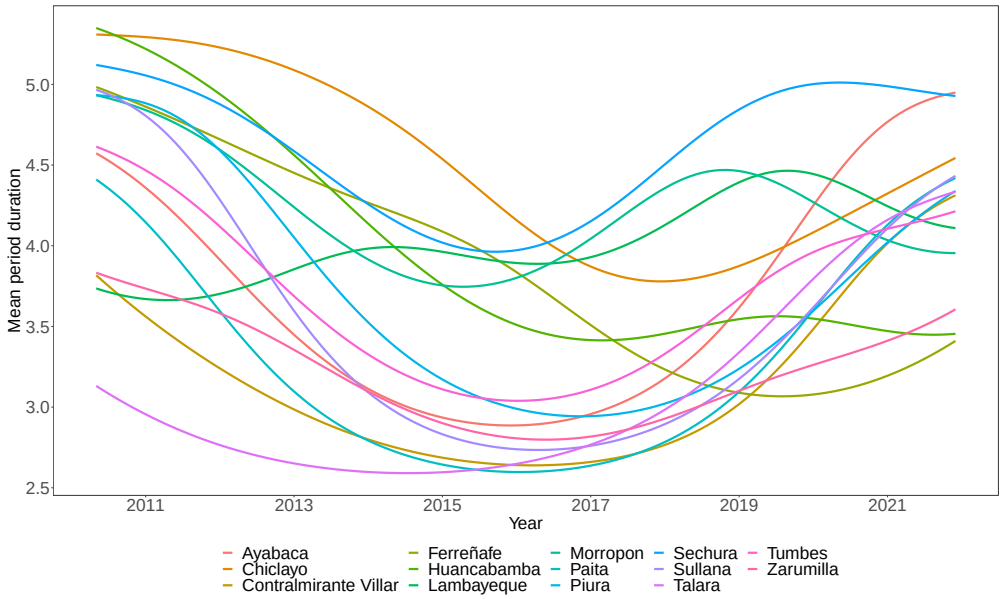


Figure SI C.7: **Mean duration of multiannual dengue cycles.** Provinces' mean duration (in years) of wavelet-based reconstructed multiannual cycles of dengue incidence rate (DIR) per 100,000 across the studied period of 2010–2021.

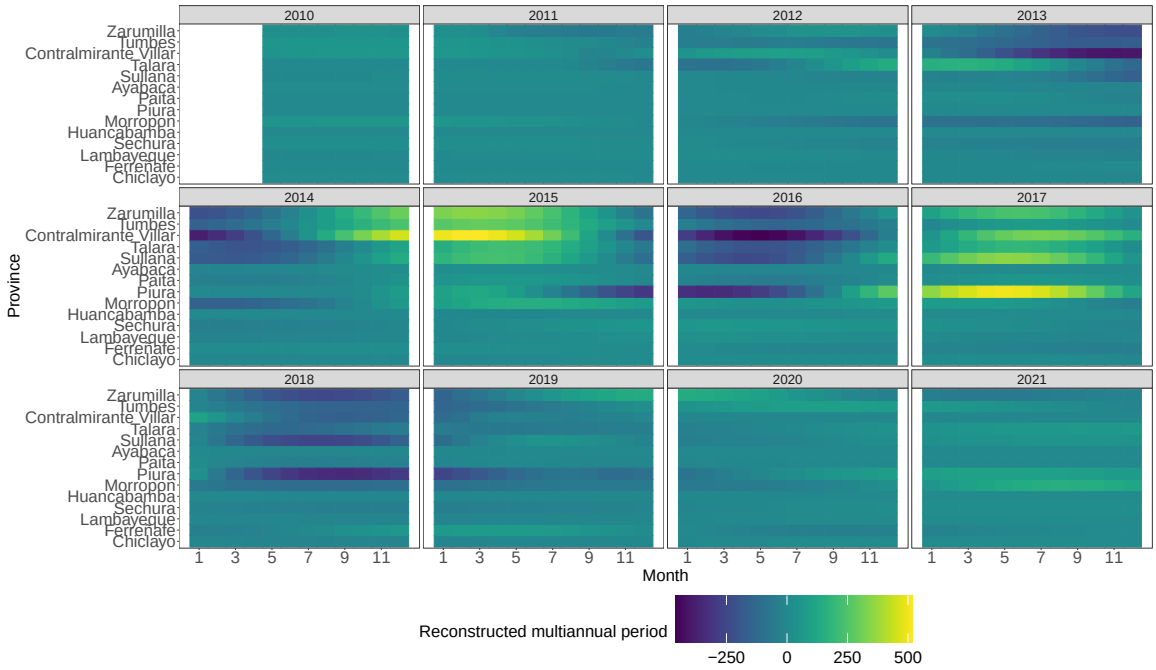


Figure SI C.8: **Wavelet reconstructions of multiannual dengue cycles in higher temporal resolution.** Multiannual cycles of province-level dengue incidence are sub-divided into individual years, where provinces are sorted by latitude in ascending order (from south to north).

Phase coherence and epidemic synchrony

Table SI C.5: **Summary statistics of provinces' phase coherence.** For both reconstructed annual and multiannual dengue cycles, for each (focal) province at each individual time point, we calculated the proportion of provinces with whom the focal province shared a statistically significant coherence. Here, we present the median, lower quartile (Q1), and upper quartile (Q3) of these proportions as a summary of the province-level coherence between the cycles across the study period.

Time Series	Median	Q1	Q3
Reconstructed Annual Cycle	0.62	0.46	0.77
Reconstructed Multiannual Cycle	0.46	0.31	0.62

Table SI C.6: **Relationship between urbanised provinces and coherence.** For both annual and multiannual dengue cycles, the correlation between the proportion of a province's cycles with statistically significant coherence and the province's percentage of population living in urban areas.

Time Series	Correlation (95% CI)
Reconstructed Annual Cycle	0.58 (0.07, 0.85)
Reconstructed Multiannual Cycle	0.46 (-0.09, 0.80)

Table SI C.7: **Summary statistics of provinces' epidemic synchrony.** For raw reported dengue incidence, reconstructed annual cycles, and reconstructed multiannual cycle, we calculated each province's median correlation (r) with the corresponding time series of the other 13 provinces. Here, we report the median, lower quartile (Q1), and upper quartile (Q3) of these correlations as a summary of the average province-level correlations between the time series.

Time Series	**Median** r	**Q1**	**Q3**
Raw Dengue Incidence Rates	0.46	0.34	0.54
Reconstructed Annual Cycle	0.76	0.69	0.80
Reconstructed Multiannual Cycle	0.60	0.28	0.66

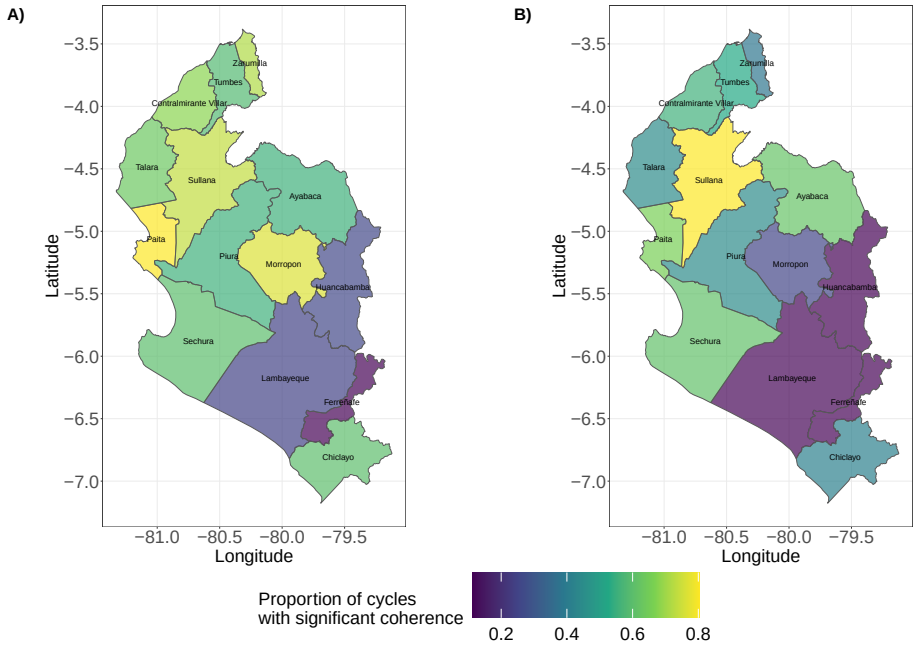


Figure SI C.9: **Spatial distribution of coherence.** Maps depict the proportion of each provinces' reconstructed **A)** annual and **B)** multiannual cycles of Dengue Incidence Rates (DIRs) which share statistically significant coherence with other provinces.

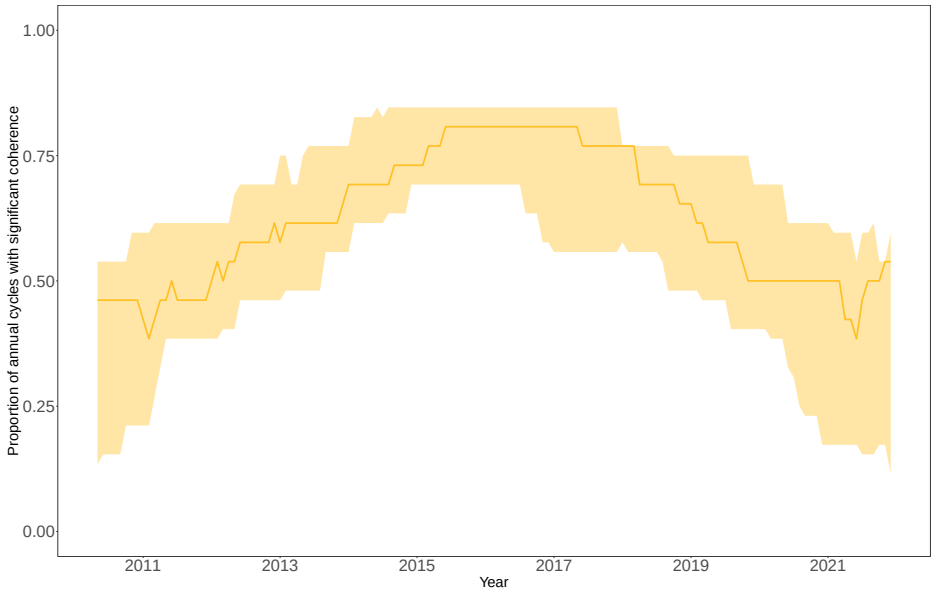


Figure SI C.10: **Temporal distribution of annual phase coherence.** For each province (the so-called focal province) at each time point, we calculated the proportion of provinces with whom the focal province shared a statistically significant coherence. Here, we summarise the distribution of these proportions over time by depicting the median proportion at each time point (gold line) and the lower and upper quartiles of proportions.

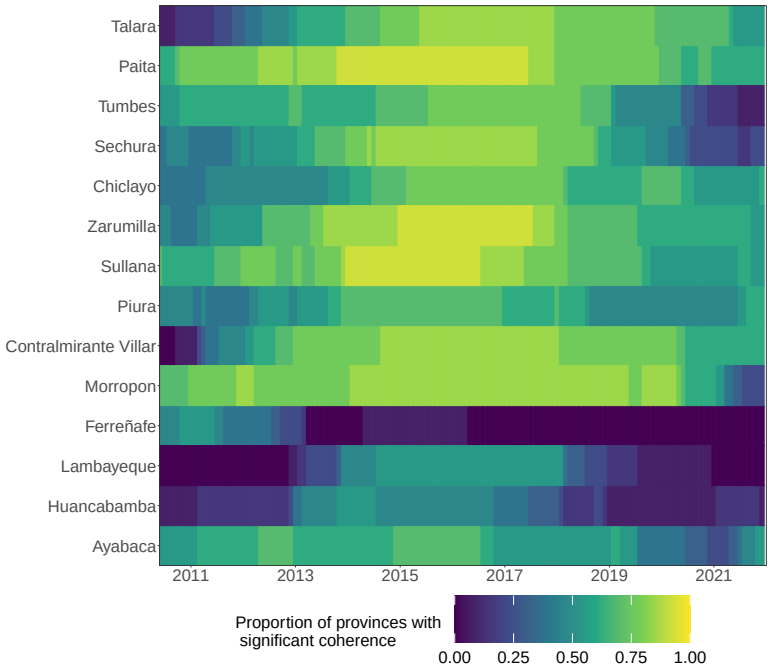


Figure SI C.11: **Coherence by level of urbanisation.** Proportion of a focal province's annual dengue cycles which share statistically significant coherence with the cycles of other provinces at each time point. Provinces are sorted in ascending order by proportion of the population that lives in an urban area.

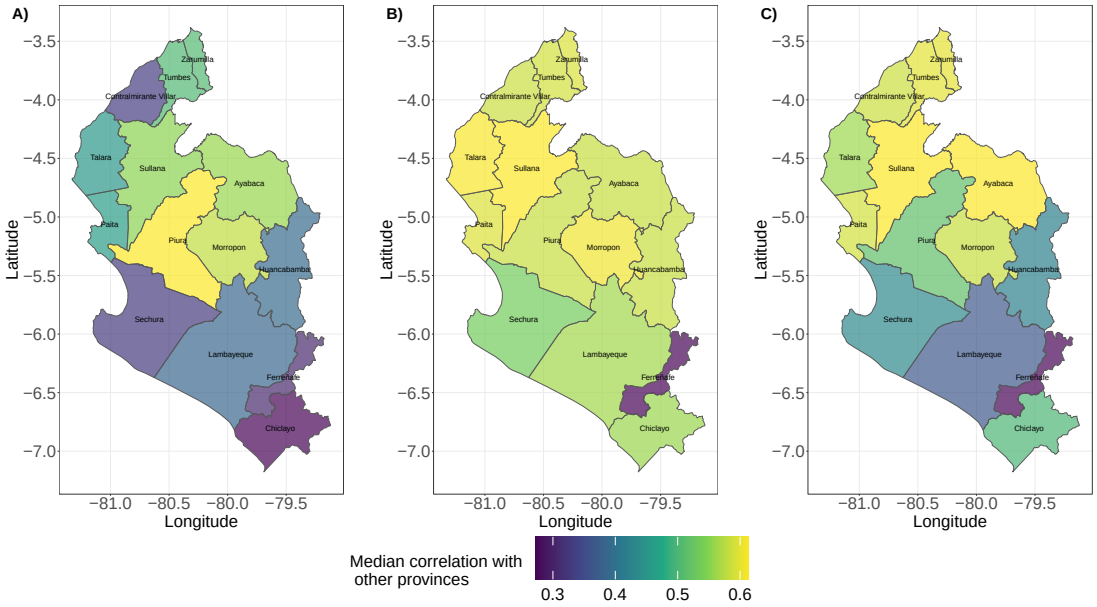


Figure SI C.12: **Spatial distribution of epidemic synchrony.** Maps depict the median correlation between a focal province's **A)** raw Dengue Incidence Rate (DIR) time series, **B)** reconstructed annual cycle or **C)** multiannual cycle, and the corresponding time series of all other provinces.

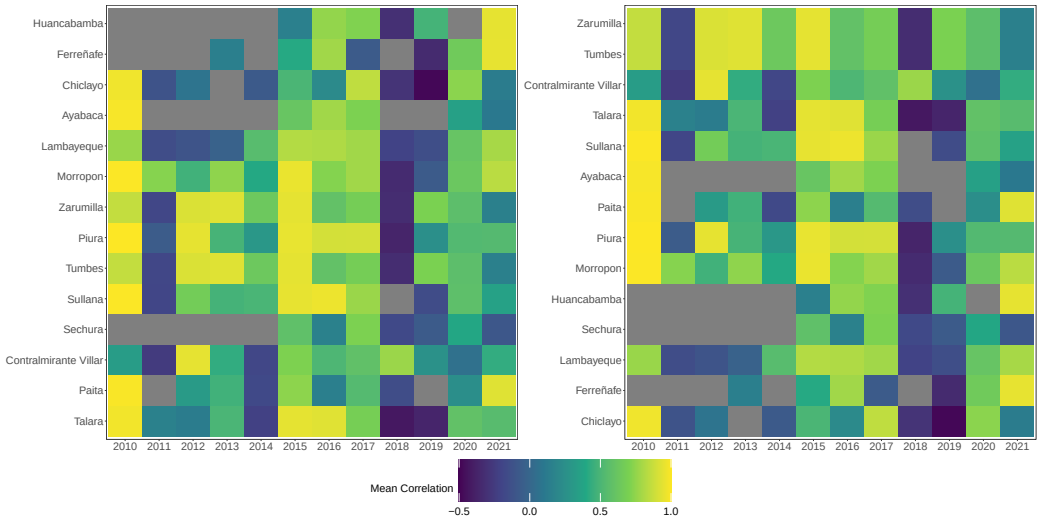


Figure SI C.13: **Epidemic synchrony across time and space.** Visualisation of the mean average correlation between a province's raw monthly Dengue Incidence Rate (DIR) time series and the other provinces. Provinces are sorted in ascending order **(A)** by longitude (from west to east) and **B** by latitude (from south to north).

Climatic, geographic, and human influences on epidemic dynamics

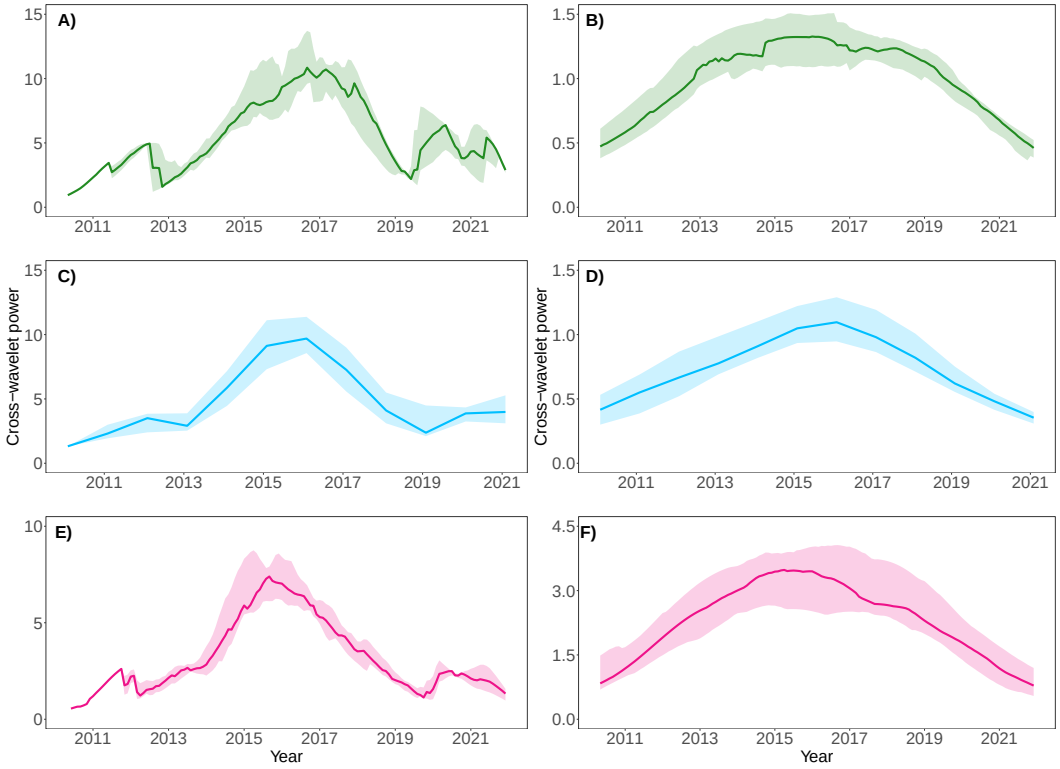


Figure SI C.14: **Distributions of cross-wavelet power between climatic conditions and reconstructed dengue cycles.** Panels summarise the distribution of cross-wavelet power between either reconstructed annual (left column) or multiannual dengue cycles (right column) and climatic conditions. **(A)-(B)**: Cross-wavelet power between reconstructed annual/multiannual cycles and monthly average maximum temperature. **(C)-(D)**: Cross-wavelet power between reconstructed annual/multiannual cycles and monthly total precipitation. **(E)-(F)**: Cross-wavelet power between reconstructed annual/multiannual cycles and monthly El Niño Coastal Index. Solid lines depict the median cross-wavelet power at each time, whilst the shaded regions depict the interquartile range – IQR; ranging from the first/lowest quartile (Q1) to the third/upper quartile (Q3).

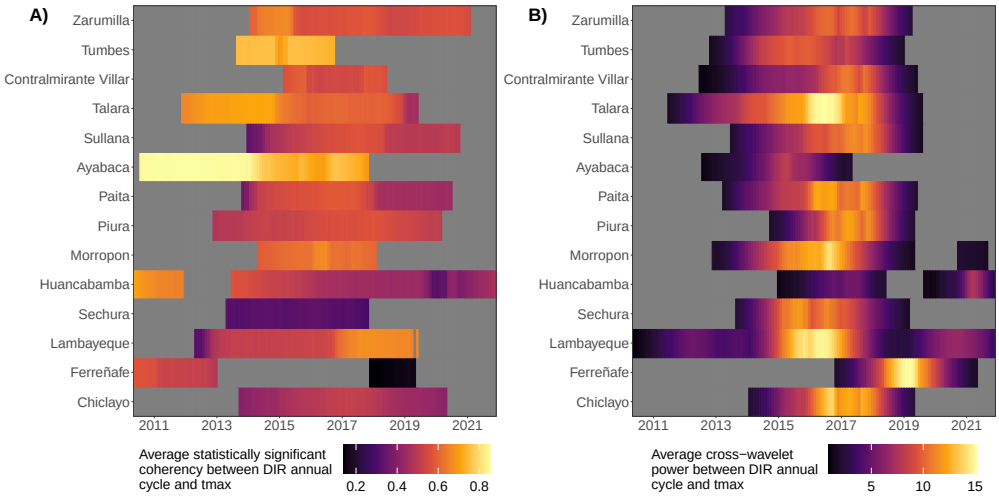


Figure SI C.15: **Maximum temperature coherence and cross-wavelet power (annual cycles). (A)**: Provinces' average statistically significant coherence between reconstructed annual dengue cycles and monthly average daily maximum temperature (denoted tmax above). **(B)**: Provinces' average cross-wavelet power between reconstructed annual dengue cycles and monthly average daily maximum temperature. Provinces are sorted in ascending order by latitude (from south to north), and non-filled (grey) cells depict non-significant wavelet coherency/cross-wavelet power.

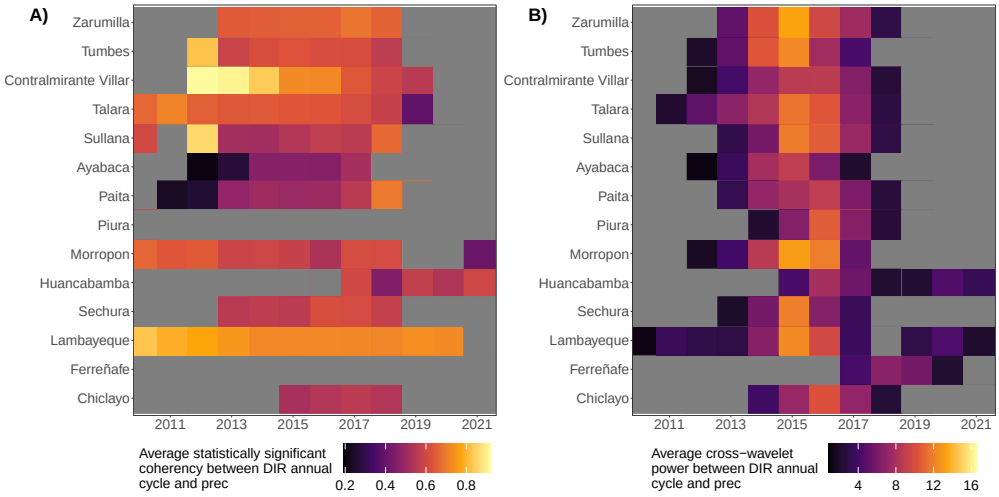


Figure SI C.16: **Monthly precipitation coherence and cross-wavelet power (annual cycles). (A)**: Provinces' average statistically significant coherence between reconstructed annual dengue cycles and monthly total precipitation (denoted prec above). **(B)**: Provinces' average cross-wavelet power between reconstructed annual dengue cycles and monthly total precipitation. Provinces are sorted in ascending order by latitude (from south to north), and non-filled (grey) cells depict non-significant wavelet coherency/cross-wavelet power.

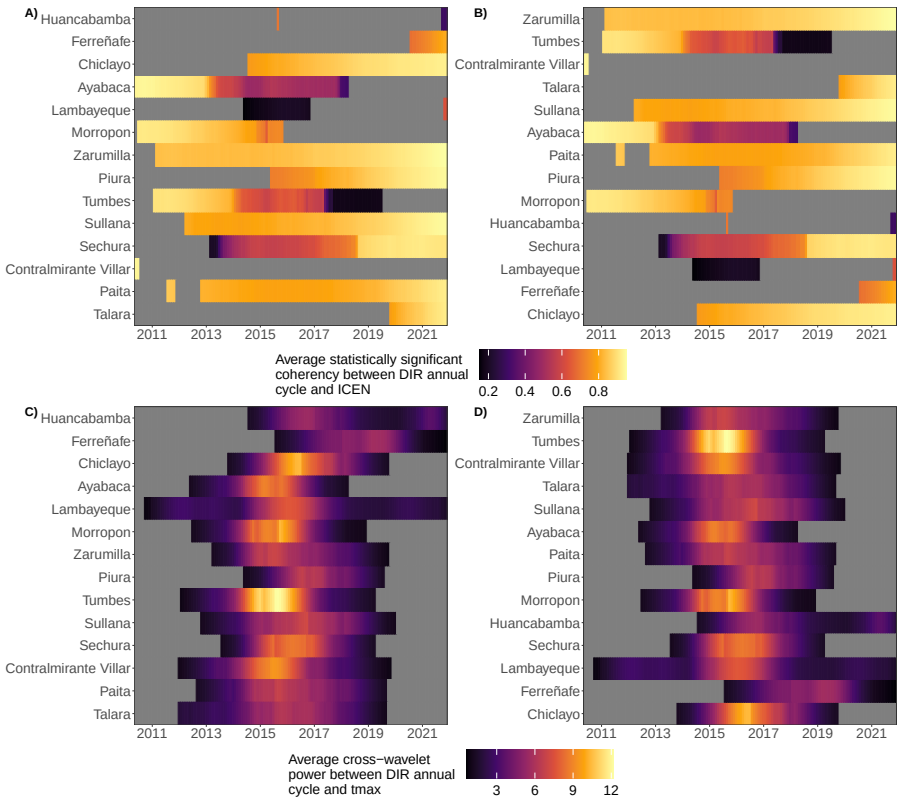


Figure SI C.17: **El Niño Coastal Index (ICEN) coherence and cross-wavelet power (annual and multiannual cycles).** **(A)** and **B**: Provinces' average statistically significant coherence between reconstructed annual dengue cycles and monthly El Niño Coastal Index (denoted ICEN above), where provinces are arranged in ascending order by longitude (from west to east) and latitude (from south to north) respectively. **(C)** and **D**: Provinces' average cross-wavelet power between reconstructed multiannual dengue cycles and monthly El Niño Coastal Index, where provinces are again arranged in ascending order by longitude and latitude respectively. Non-filled (grey) cells depict non-significant wavelet coherency/cross-wavelet power.

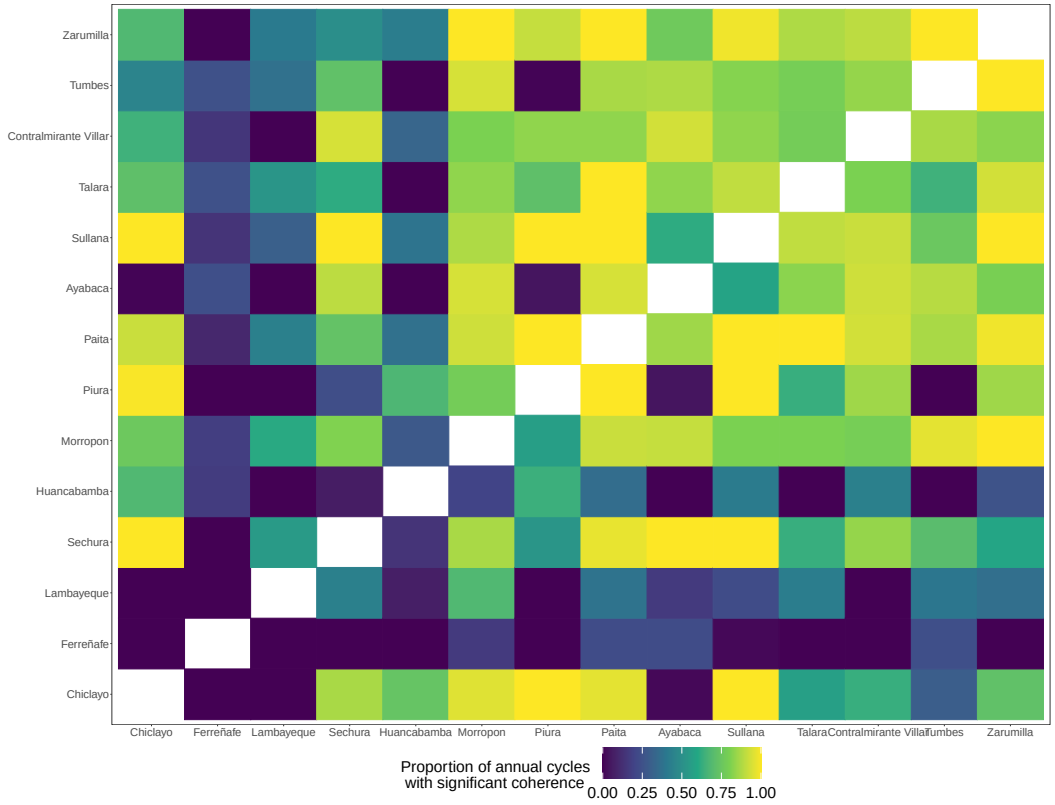


Figure SI C.18: **Pairwise coherence visualisation.** Cells depict proportion of annual cycles between provinces that are statistically significant, where provinces are sorted by latitude (from south to north). White filled cells denote the diagonal entries of a focal province's coherence with its own time series. The visualisation captures the strength of relationships between a province and other provinces, ranging from close neighbours to provinces further away.

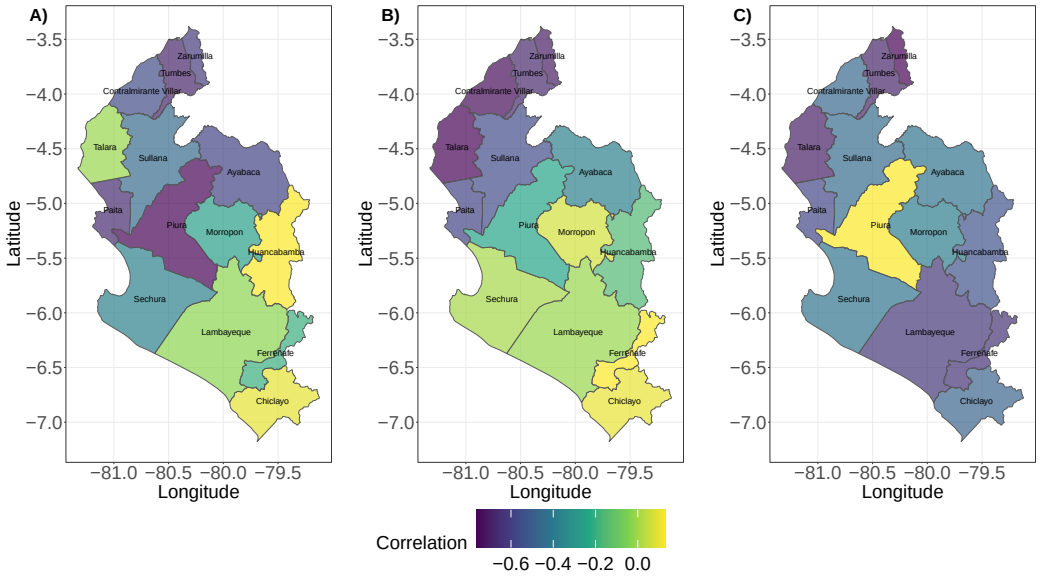


Figure SI C.19: **Relationship between provinces' pairwise distance and epidemic dynamics.** **(A)**: Correlation between provinces' pairwise distances and the corresponding median pairwise Pearson correlation of Dengue Incidence Rate (DIR) time series. **(B)**: Correlation between provinces' pairwise distances and the proportion of the reconstructed annual dengue cycles sharing significant coherence. **(C)**: Correlation between provinces' pairwise distance and the proportion of the reconstructed multiannual dengue cycles sharing significant coherence.

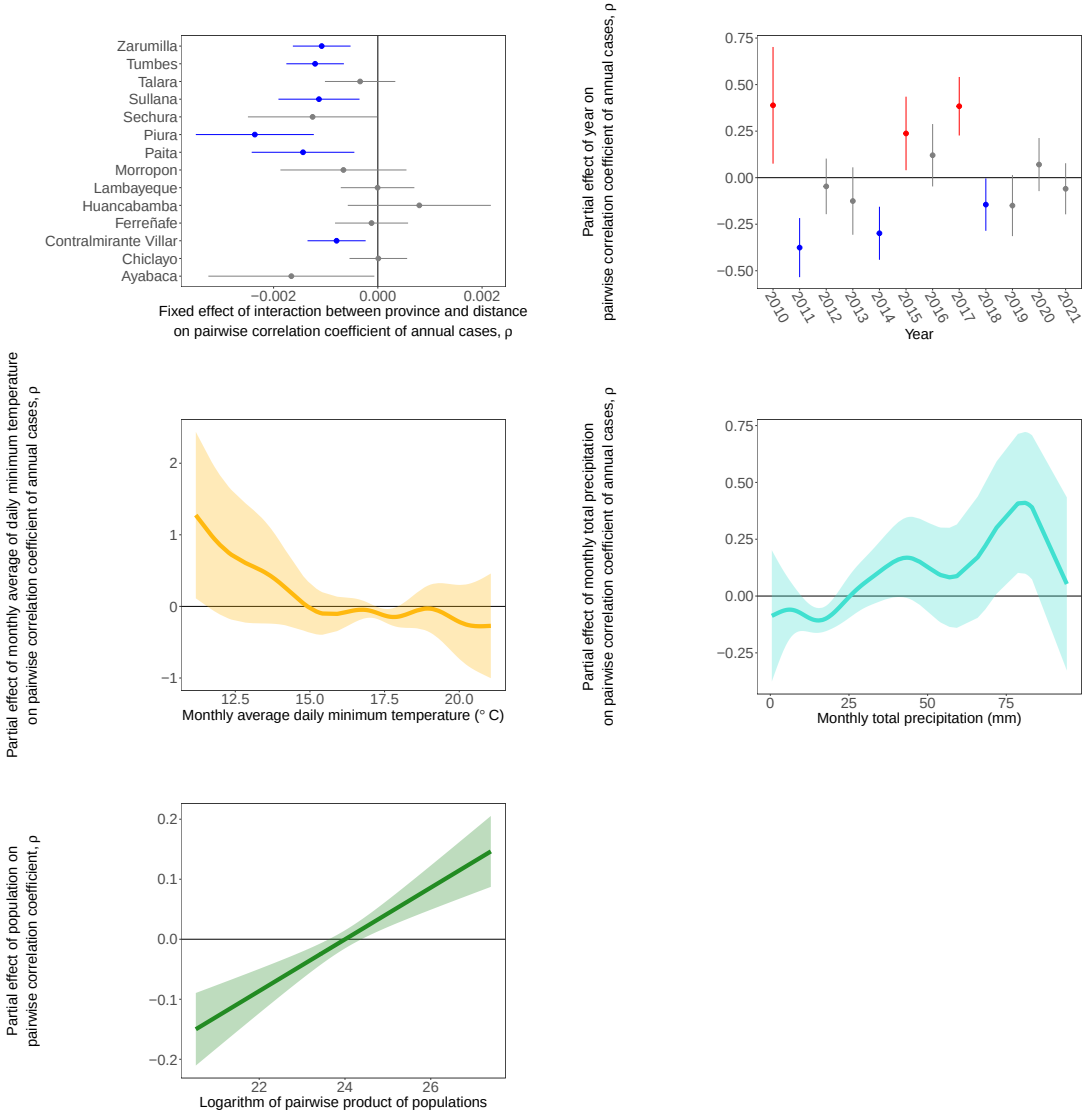


Figure SI C.20: **Results from best-fitting model of epidemic synchrony.** Top panel: Visualisation of fixed effect of interaction between and distance and partial effect of year on the pairwise annual correlation coefficient (between annual case cycles), where blue and red denote negative and positive statistically significant estimated effects respectively. Middle panel: Visualisation of the partial effects of minimum temperature (left) and precipitation (right) on the pairwise annual correlation coefficient, where the x-axes denotes the yearly average of the corresponding monthly averages. Bottom panel: Visualisation of the partial effect of the pairwise product of logarithm of populations on the pairwise annual correlation coefficient. Shaded regions depict the 95% confidence intervals in each case.

C.9.2 Climate-based modelling

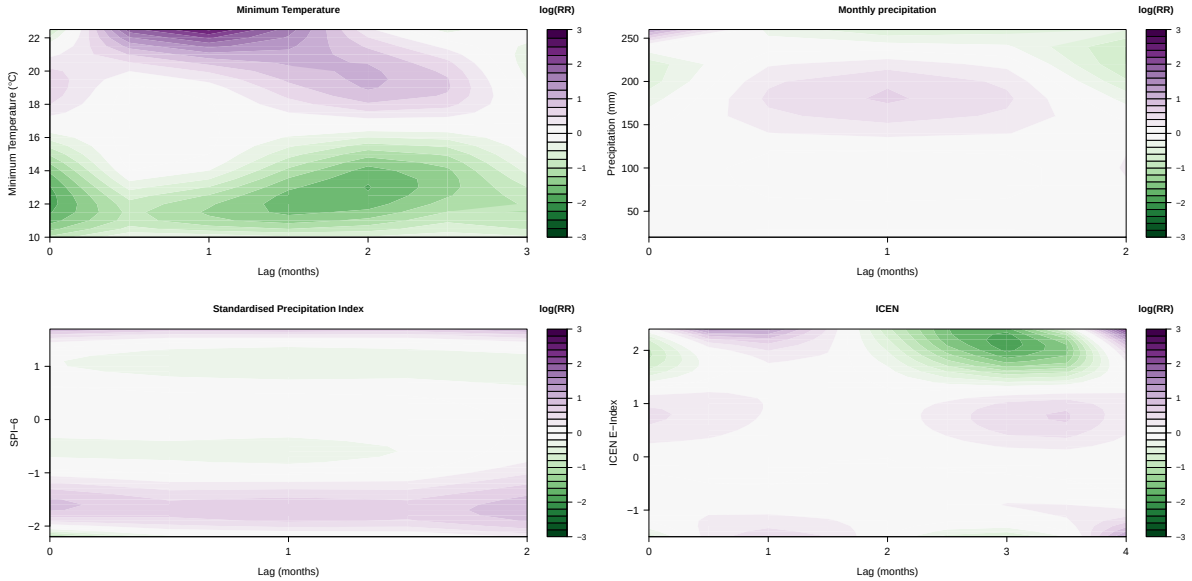


Figure SI C.21: **Model-based exposure-lag-response relationships between climatic variables and dengue incidence.** Across 2010 to 2021, the relative risk (RR) of dengue incidence, on a logarithm scale, for each of our climatic DLNMs (Distributed Lag Nonlinear Models) in our climate-based Bayesian spatiotemporal model. RR was defined relative to the risk induced by the mean observed value of each climate variable, and thus, log RR values above 0 (pink to purple) correspond to elevated RR of dengue incidence, whilst values below 0 (green) correspond to reduced RR. The DLNMs were included in a model alongside various temporal, spatial, and spatiotemporal effects, as well as fixed effects.

C.9.3 Probabilistic ensemble forecasting

The following section presents results and tables for probabilistic forecasting of dengue cases across i) the model development period (2010 to 2017 inclusive) and ii) the testing period (2018 to 2021 inclusive). All probabilistic forecasts were made with a forecast horizon of one month and excluded all future data from model training.

Model development period – Summary

Table SI C.8: **Results for forecasting log(Cases + 1) in model development period.** In these model rankings, we summarise proper scores, predictive performance, calibration, and outbreak detection capabilities of forecasting models. Models are arranged by mean weighted interval score (WIS), from best (lowest) to worst (highest). Bias measures the bias of each models' predictive quantiles, PI coverage measures the coverage of the 95% prediction intervals (PIs), R^2 denotes the proportion of variation in log(Cases + 1) explained by the models' posterior median estimates, and the corresponding root mean square error (RMSE) for log(Cases + 1) by using the posterior median. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk. As some forecasting models require a minimum number of observations, these comparisons are from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

	Model	WIS	Bias	PI Coverage	R^2	RMSE
1	TimeGPT	0.41	-0.04	93.2%	0.84	0.86
2	Median-NoBayes *	0.45	-0.16	94.4%	0.82	0.95
3	Median *	0.46	-0.22	93.9%	0.82	0.97
4	Ew-Mean-NoBayes *	0.47	-0.13	90.1%	0.82	0.96
5	Median-NoBase *	0.49	-0.25	95.5%	0.80	1.04
6	Median-NoCov *	0.49	-0.10	90.6%	0.79	1.01
7	SARIMA	0.50	-0.04	87.5%	0.79	1.02
8	EW-Mean *	0.51	-0.19	89.5%	0.79	1.08
9	EW-Mean-NoCov *	0.52	-0.16	88.7%	0.79	1.06
10	EW-Mean-NoBase *	0.55	-0.21	90.9%	0.76	1.18
11	Baseline	0.57	0.01	78.2%	0.79	1.00
12	TCN	0.59	-0.22	87.3%	0.75	1.15
13	TimeGPT-NoCov	0.89	-0.32	88.7%	0.46	1.95
14	Bayes-Climate	1.03	-0.41	76.3%	0.28	2.39

Model development period – Individual forecasting models

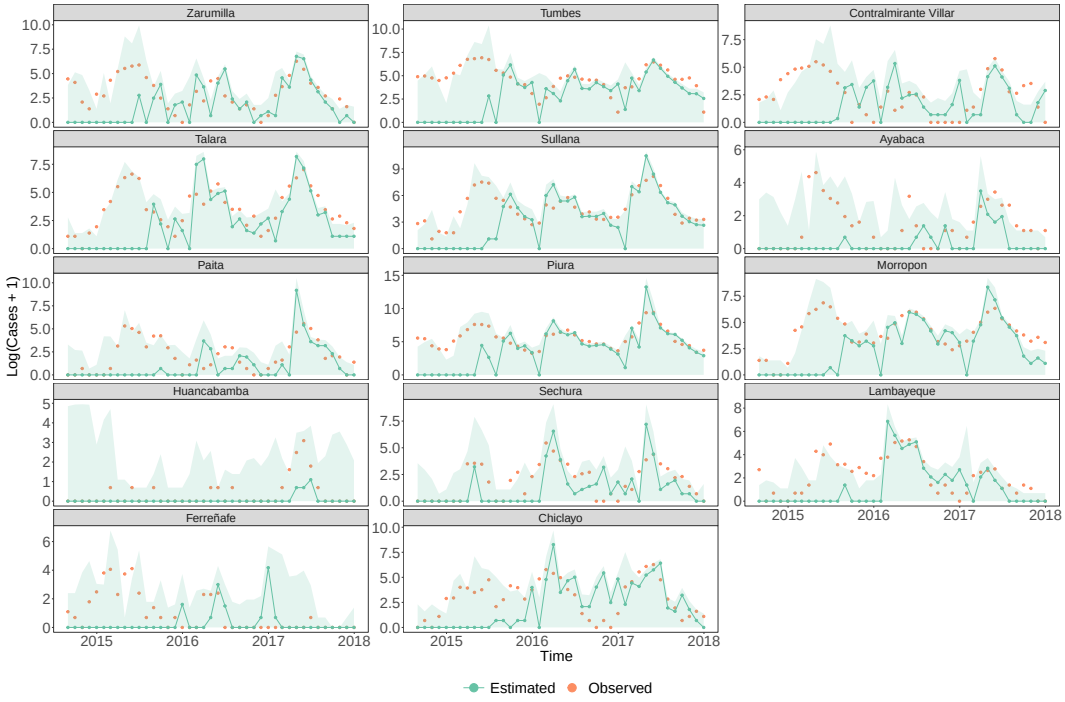


Figure SI C.22: **Bayesian climate model's predictive performance across the model development period.** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green). The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics. As some forecasting models require a minimum number of observations, for equitable comparison across models for the model development period, we present results from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

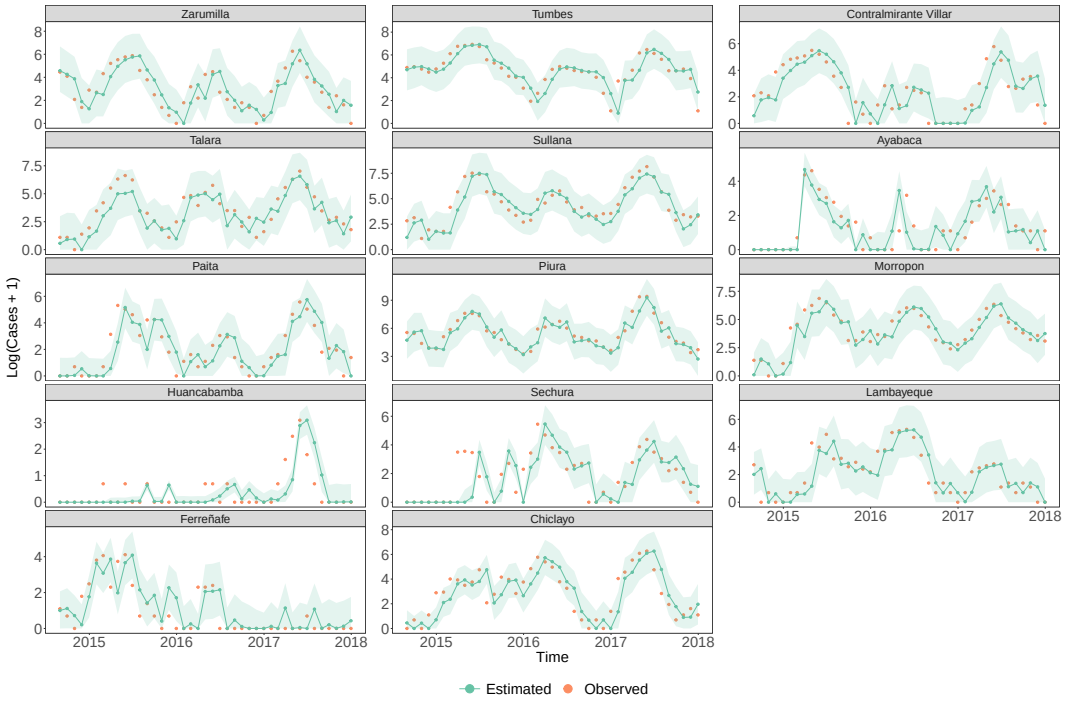


Figure SI C.23: **SARIMA models' predictive performance across the model development period.** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green). The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics. As some forecasting models require a minimum number of observations, for equitable comparison across models for the model development period, we present results from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

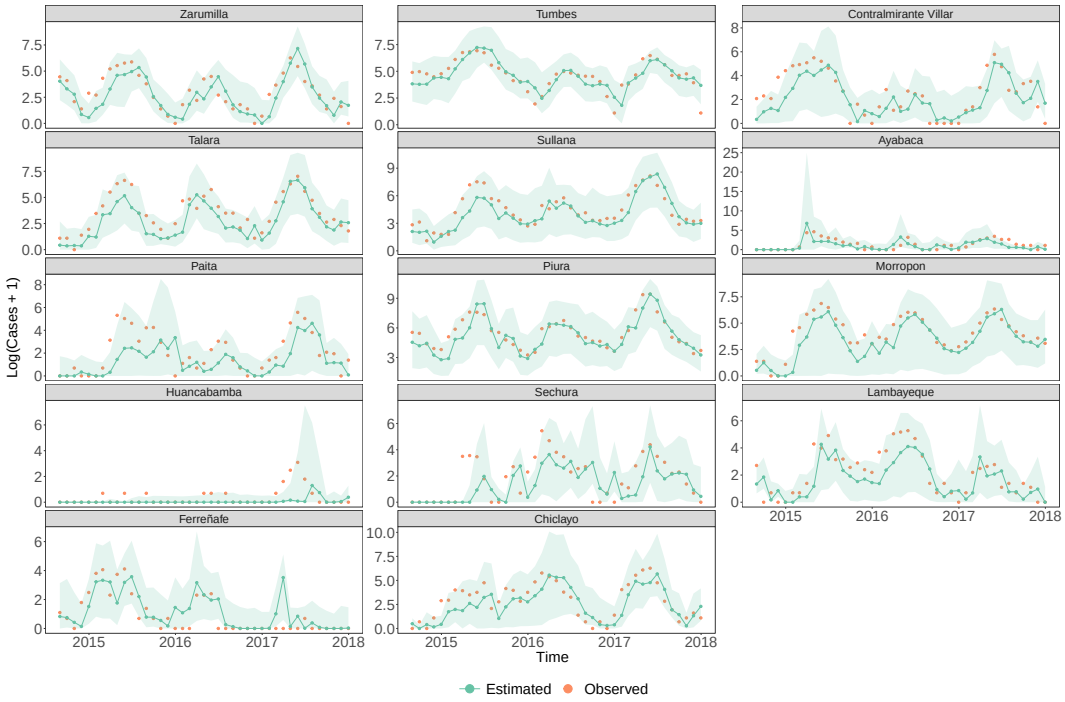


Figure SI C.24: **TCN models' predictive performance across the model development period.** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green). The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics. As some forecasting models require a minimum number of observations, for equitable comparison across models for the model development period, we present results from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

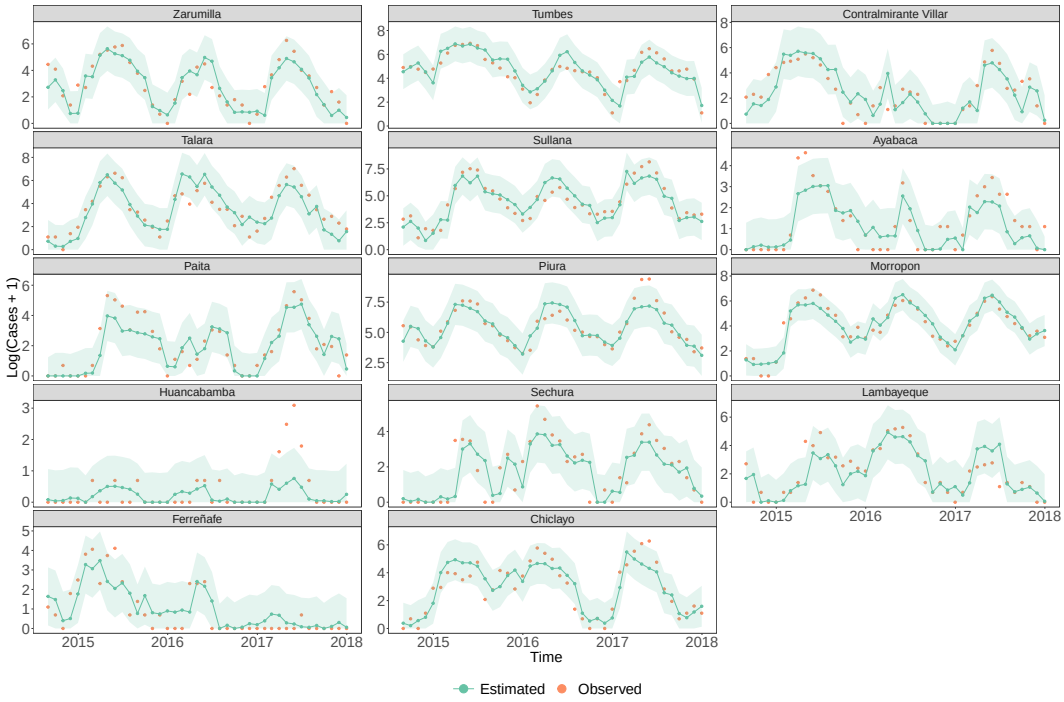


Figure SI C.25: **TimeGPT model's predictive performance across the model development period.** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green). The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics. As some forecasting models require a minimum number of observations, for equitable comparison across models for the model development period, we present results from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

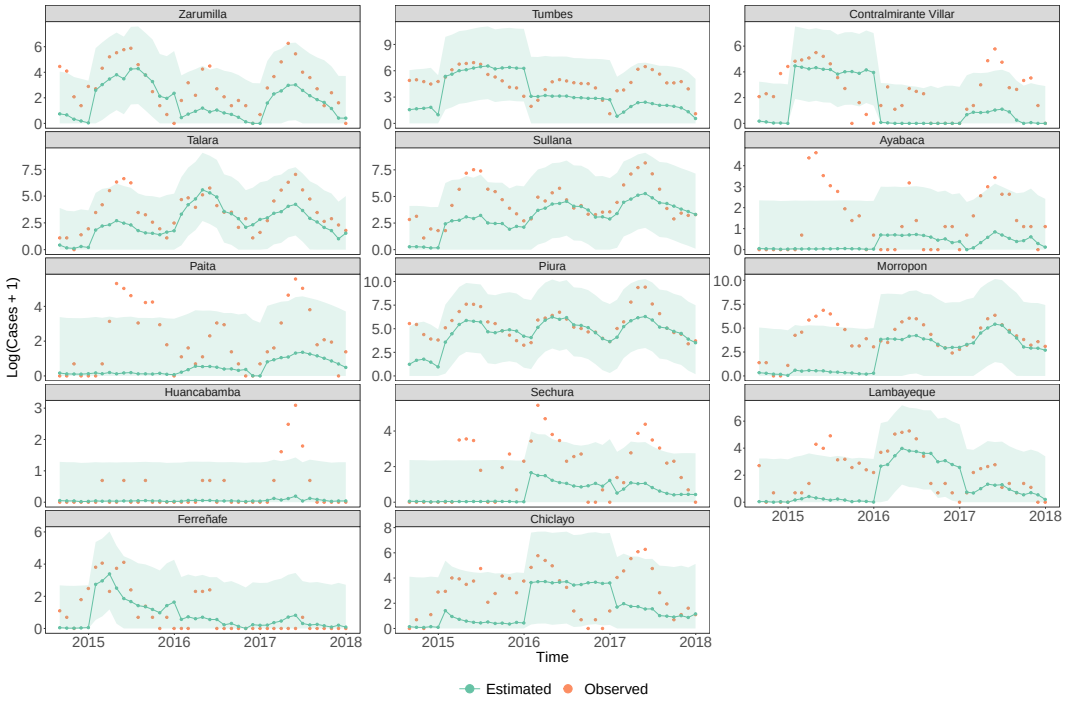


Figure SI C.26: **Predictive performance of TimeGPT model without covariates across the model development period.** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green). The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics. As some forecasting models require a minimum number of observations, for equitable comparison across models for the model development period, we present results from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

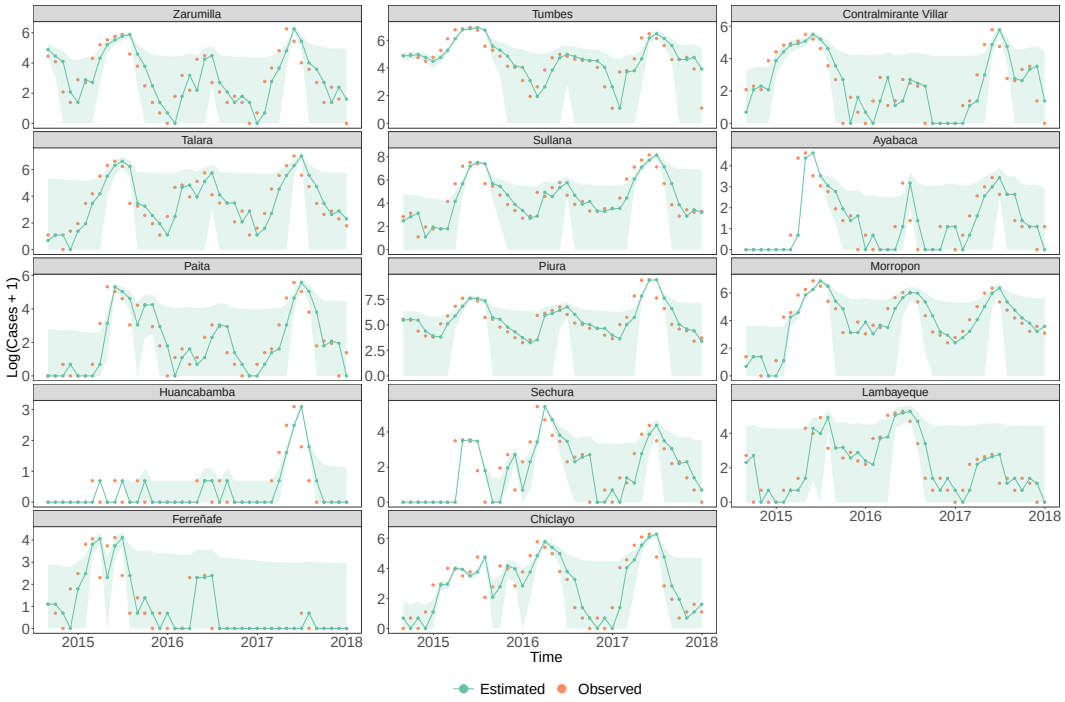


Figure SI C.27: **Baseline model's predictive performance across the model development period.** We visualise results obtained for the epidemiologically naive baseline forecaster. For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green). The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics. As some forecasting models require a minimum number of observations, for equitable comparison across models for the model development period, we present results from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

Model development period — Coverage

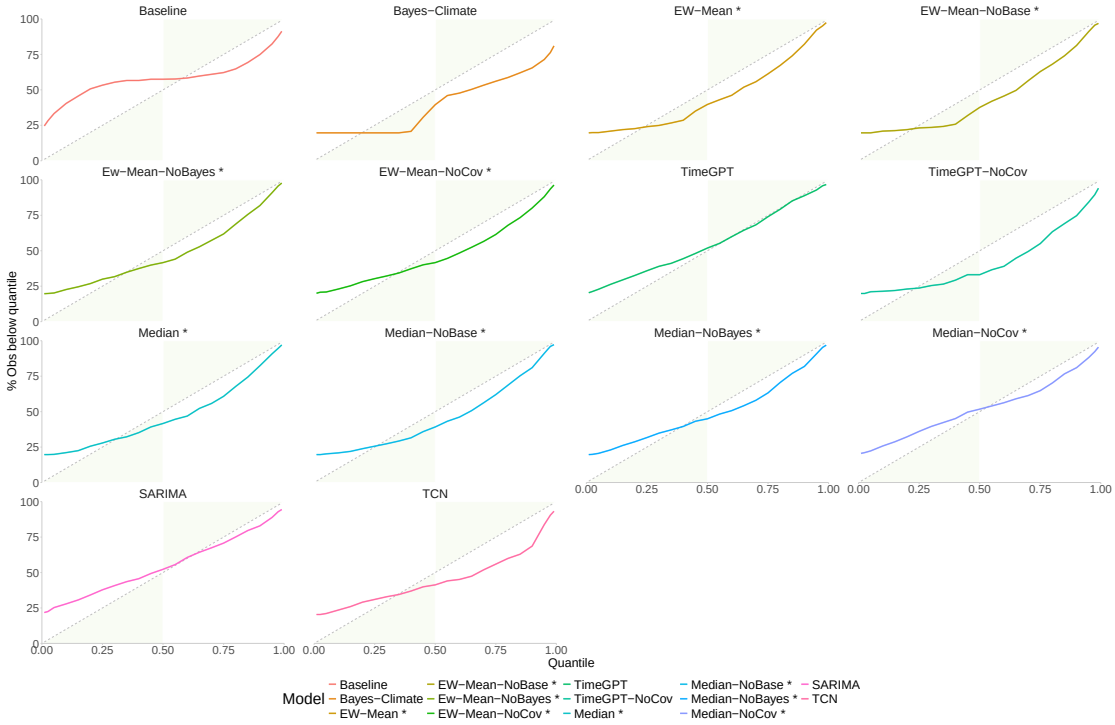


Figure SI C.28: **Quantile coverage for quantile-based forecasting models across the model development period.** For each forecasting model, the quantile coverage is plotted where the y-axis denotes the percentage of observed values (log(Cases + 1)) that lie below the corresponding model-based predictive quantile level shown on the x-axis. Quantile coverage can be interpreted as a measure of probabilistic calibration i.e. consistency of a model's probabilities with those actually observed. The green shaded polygons denote areas of conservative behaviour i.e. empirical coverage outweighs the nominal coverage required. Details of the model abbreviations are provided in Table SI C.2. As some forecasting models require a minimum number of observations, these comparisons are from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

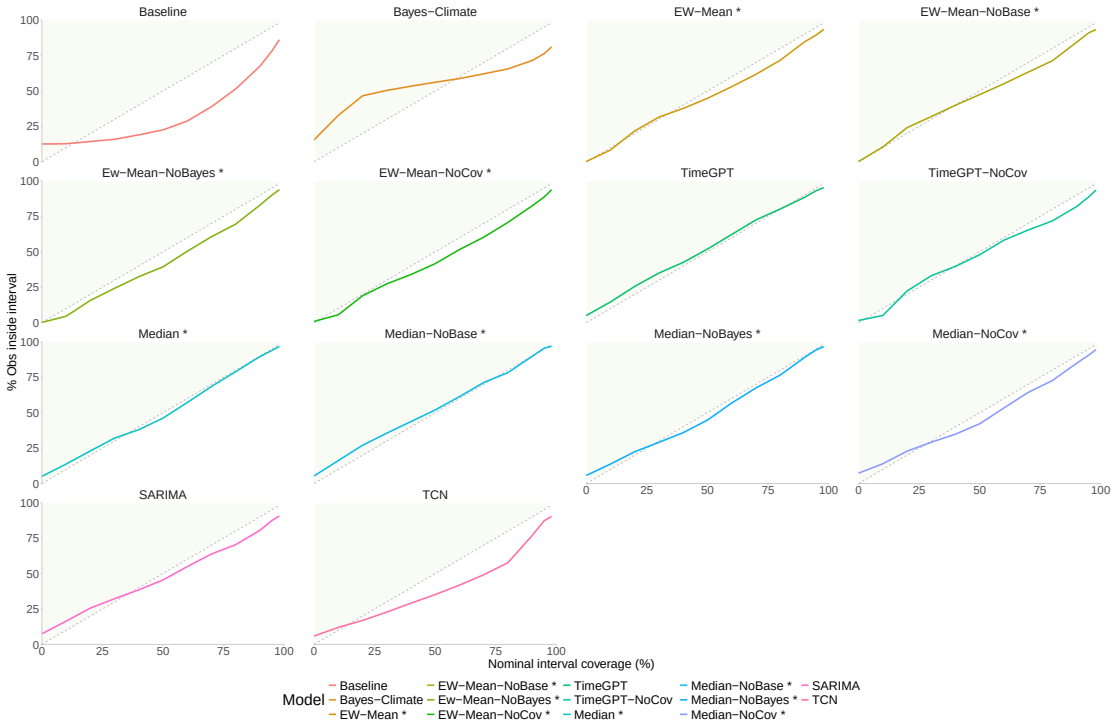


Figure SI C.29: **Interval coverage for quantile-based forecasting models across the model development period.** For each forecasting model, the interval coverage is plotted where the y-axis denotes the percentage of observed values ((log(Cases + 1))) that lie below the corresponding model-based predictive quantile level shown on the x-axis. Interval coverage can be interpreted as a measure of probabilistic calibration i.e. consistency of a model's probabilities with those actually observed. The green shaded polygons denote areas of conservative behaviour i.e. empirical coverage outweighs the nominal coverage required. Interval coverage alone is not a measure of forecasting skill, as it only focuses on probabilistic calibration, and thus, we also report values of a proper scoring rule, the weighted interval score (WIS) which measures calibration and sharpness simultaneously. Details of the model abbreviations are provided in Table SI C.2. As some forecasting models require a minimum number of observations, these comparisons are from the first month at which all forecasting models can make a prediction (August 2014) until the end of the model development period (December 2017).

Testing period – Summary tables

Table SI C.9: **Results for forecasting log(Cases + 1) in the testing period of 2018 to 2021.** We summarise proper scores, predictive performance, calibration, and outbreak detection capabilities of forecasting models. Models are arranged by mean weighted interval score (WIS), from best (lowest) to worst (highest). Bias measures the bias of each models' predictive quantiles, PI coverage measures the coverage of the 95% prediction intervals (PIs), R^2 denotes the proportion of variation in log(Cases + 1) explained by the models' posterior median estimates, and the corresponding root mean squared error (RMSE) for log(Cases + 1) by using the posterior median. For detection of outbreak onset, *Sensitivity*, *Specificity*, and *AUC* represent the True Positive rate, True Negative rate, and Area Under the Curve in terms of forecasting skill for correctly classifying the first possible future outbreaks of DIR $\geq$ 150 per 100,000. Table 4.3 is the analogous table which depicts outbreak detection for DIR first surpassing a threshold of 50 per 100,000. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk.

	Model	WIS	Bias	PI Coverage	R^2	RMSE	Sensitivity	Specificity	AUC
1	Median *	0.34	0.03	95.2%	0.74	0.53	80.0%	94.9%	0.87
2	Median-NoBase *	0.35	0.02	94.8%	0.73	0.54	80.0%	95.3%	0.88
3	Median-NoBayes *	0.35	0.06	94.8%	0.74	0.54	80.0%	94.5%	0.87
4	EW-Mean *	0.35	0.17	87.8%	0.75	0.54	80.0%	93.7%	0.87
5	Median-NoCov *	0.36	0.04	95.2%	0.73	0.55	80.0%	93.9%	0.87
6	Ew-Mean-NoBayes *	0.36	0.21	87.5%	0.74	0.55	60.0%	95.0%	0.78
7	EW-Mean-NoCov *	0.36	0.14	91.2%	0.74	0.55	60.0%	90.4%	0.75
8	Trained *	0.36	0.07	92.4%	0.72	0.56	80.0%	93.4%	0.87
9	EW-Mean-NoBase *	0.36	0.16	87.1%	0.74	0.55	60.0%	97.3%	0.79
10	SARIMA	0.36	0.03	93.5%	0.73	0.56	80.0%	94.2%	0.87
11	Prov-Trained *	0.37	0.05	91.1%	0.72	0.56	80.0%	94.7%	0.87
12	Baseline	0.38	0.04	97.3%	0.74	0.52	60.0%	55.0%	0.57
13	TCN	0.39	0.11	91.2%	0.70	0.59	80.0%	94.1%	0.87
14	TimeGPT	0.41	0.08	88.1%	0.69	0.62	60.0%	94.1%	0.77
15	TimeGPT-NoCov	0.42	0.12	85.6%	0.67	0.62	40.0%	95.3%	0.68
16	Bayes-Climate	0.52	-0.20	87.6%	0.58	0.71	40.0%	88.6%	0.64

Table SI C.10: **Detection of all outbreaks with DIR ≥ 50 per 100,000 across all months of 2018 to 2021.** For detection of any outbreak months, *Sensitivity*, *Specificity*, and *AUC* represent the True Positive rate, True Negative rate, and Area Under the Curve in terms of forecasting skill for correctly classifying future outbreaks of DIR ≥ 50 per 100,000. Table SI C.11 is the analogous table which depicts outbreak detection for DIR reaching a threshold of 150 per 100,000. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk. Models are presented in the same ordering as Table 4.3 to enable comparison with the models' performance across other performance metrics.

	Model	Sensitivity	Specificity	AUC
1	Median *	91.7%	95.2%	0.93
2	Median-NoBase *	89.6%	95.4%	0.92
3	Median-NoBayes *	91.7%	94.7%	0.93
4	EW-Mean *	89.6%	96.5%	0.93
5	Median-NoCov *	91.7%	92.5%	0.92
6	Ew-Mean-NoBayes *	89.6%	96.8%	0.93
7	EW-Mean-NoCov *	85.4%	96.3%	0.91
8	Trained *	91.7%	93.6%	0.93
9	EW-Mean-NoBase *	89.6%	96.3%	0.93
10	SARIMA	89.6%	95.8%	0.93
11	Prov-Trained *	89.6%	95.8%	0.93
12	Baseline	81.2%	96.2%	0.89
13	TCN	89.6%	96.3%	0.93
14	TimeGPT	79.2%	95.0%	0.87
15	TimeGPT-NoCov	87.5%	88.0%	0.88
16	Bayes-Climate	64.6%	94.1%	0.79

Table SI C.11: **Detection of all outbreaks with DIR ≥ 150 per 100,000 across all months of 2018 to 2021.** For detection of any outbreak months, *Sensitivity*, *Specificity*, and *AUC* represent the True Positive rate, True Negative rate, and Area Under the Curve in terms of forecasting skill for correctly classifying future outbreaks of DIR ≥ 150 per 100,000. Table SI C.10 is the analogous table which depicts outbreak detection for DIR reaching a threshold of 50 per 100,000. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk. Models are presented in the same ordering as Table 4.3 to enable comparison with the models' performance across other performance metrics.

	Model	Sensitivity	Specificity	AUC
1	Median *	85.7%	95.3%	0.91
2	Median-NoBase *	85.7%	95.0%	0.90
3	Median-NoBayes *	85.7%	97.1%	0.91
4	EW-Mean *	85.7%	96.4%	0.91
5	Median-NoCov *	85.7%	93.2%	0.89
6	Ew-Mean-NoBayes *	85.7%	96.2%	0.91
7	EW-Mean-NoCov *	78.6%	96.8%	0.88
8	Trained *	85.7%	95.9%	0.91
9	EW-Mean-NoBase *	85.7%	95.9%	0.91
10	SARIMA	78.6%	96.0%	0.87
11	Prov-Trained *	85.7%	94.8%	0.90
12	Baseline	64.3%	98.5%	0.81
13	TCN	85.7%	96.2%	0.91
14	TimeGPT	78.6%	95.1%	0.87
15	TimeGPT-NoCov	71.4%	92.9%	0.82
16	Bayes-Climate	71.4%	93.0%	0.82

Table SI C.12: **Results for forecasting DIR in the testing period of 2018 to 2021.** We summarise proper scores, predictive performance, calibration, and outbreak detection capabilities of forecasting models. Models are arranged by mean weighted interval score (WIS), from best (lowest) to worst (highest). Bias measures the bias of each models' predictive quantiles, PI coverage measures the coverage of the 95% prediction intervals (PIs), R^2 denotes the proportion of variation in DIR explained by the models' posterior median estimates, and the corresponding root mean square error (RMSE) for log(Cases + 1) by using the posterior median. For detection of outbreak onset, *Sensitivity*, *Specificity*, and *AUC* represent the True Positive rate, True Negative rate, and Area Under the Curve in terms of forecasting skill for correctly classifying the first possible future outbreaks of DIR $\geq$ 50 per 100,000. Table 4.3 is the analogous table which depicts these model rankings for the logarithm of cases. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk.

	Model	WIS	Bias	PI Coverage	R^2	RMSE	Sensitivity	Specificity	AUC
1	Median *	4.95	0.04	95.2%	0.68	21.97	81.8%	94.3%	0.88
2	Median-NoBayes *	5.01	0.05	94.9%	0.68	22.13	81.8%	87.2%	0.85
3	Median-NoCov *	5.11	0.04	95.4%	0.65	22.99	81.8%	91.0%	0.86
4	Ew-Mean-NoBayes *	5.16	0.25	88.1%	0.69	21.59	63.6%	97.3%	0.80
5	Median-NoBase *	5.18	0.02	94.8%	0.67	22.43	81.8%	95.7%	0.89
6	EW-Mean *	5.19	0.24	88.7%	0.68	22.79	72.7%	95.9%	0.84
7	EW-Mean-NoCov *	5.25	0.17	91.4%	0.64	23.26	81.8%	86.5%	0.84
8	EW-Mean-NoBase *	5.36	0.23	87.5%	0.66	23.64	81.8%	94.7%	0.88
9	Trained *	5.36	0.12	92.1%	0.67	22.57	81.8%	95.4%	0.89
10	SARIMA	5.40	0.03	93.5%	0.67	23.27	63.6%	96.1%	0.80
11	Prov-Trained *	5.49	0.09	90.6%	0.66	24.28	72.7%	89.9%	0.81
12	TCN	5.67	0.11	91.2%	0.59	27.84	72.7%	96.6%	0.85
13	Baseline	6.17	0.01	97.3%	0.64	24.13	90.9%	45.0%	0.68
14	TimeGPT-NoCov	6.39	0.12	85.6%	0.47	28.32	72.7%	91.5%	0.82
15	TimeGPT	7.26	0.08	88.1%	0.54	31.29	90.9%	91.7%	0.91
16	Bayes-Climate	8.27	-0.22	87.1%	0.38	53.12	54.5%	90.1%	0.72

Table SI C.13: **Results for forecasting DIR in the testing period of 2018 to 2021.** This table is analogous to Table SI C.12, but here we report capabilities for detection of first possible future outbreaks of DIR ≥ 150 per 100,000. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk.

	Model	WIS	Bias	PI Coverage	R^2	RMSE	Sensitivity	Specificity	AUC
1	Median *	4.95	0.04	95.2%	0.68	21.97	80.0%	94.9%	0.87
2	Median-NoBayes *	5.01	0.05	94.9%	0.68	22.13	80.0%	94.5%	0.87
3	Median-NoCov *	5.11	0.04	95.4%	0.65	22.99	80.0%	93.9%	0.87
4	Ew-Mean-NoBayes *	5.16	0.25	88.1%	0.69	21.59	60.0%	95.0%	0.78
5	Median-NoBase *	5.18	0.02	94.8%	0.67	22.43	80.0%	95.3%	0.88
6	EW-Mean *	5.19	0.24	88.7%	0.68	22.79	80.0%	93.7%	0.87
7	EW-Mean-NoCov *	5.25	0.17	91.4%	0.64	23.26	60.0%	90.4%	0.75
8	EW-Mean-NoBase *	5.36	0.23	87.5%	0.66	23.64	60.0%	97.3%	0.79
9	Trained *	5.36	0.12	92.1%	0.67	22.57	80.0%	93.4%	0.87
10	SARIMA	5.40	0.03	93.5%	0.67	23.27	80.0%	94.2%	0.87
11	Prov-Trained *	5.49	0.09	90.6%	0.66	24.28	80.0%	94.7%	0.87
12	TCN	5.67	0.11	91.2%	0.59	27.84	80.0%	94.1%	0.87
13	Baseline	6.17	0.01	97.3%	0.64	24.13	60.0%	55.0%	0.57
14	TimeGPT-NoCov	6.39	0.12	85.6%	0.47	28.32	40.0%	95.3%	0.68
15	TimeGPT	7.26	0.08	88.1%	0.54	31.29	60.0%	94.1%	0.77
16	Bayes-Climate	8.27	-0.22	87.1%	0.38	53.12	40.0%	88.6%	0.64

Testing period – Individual forecasters

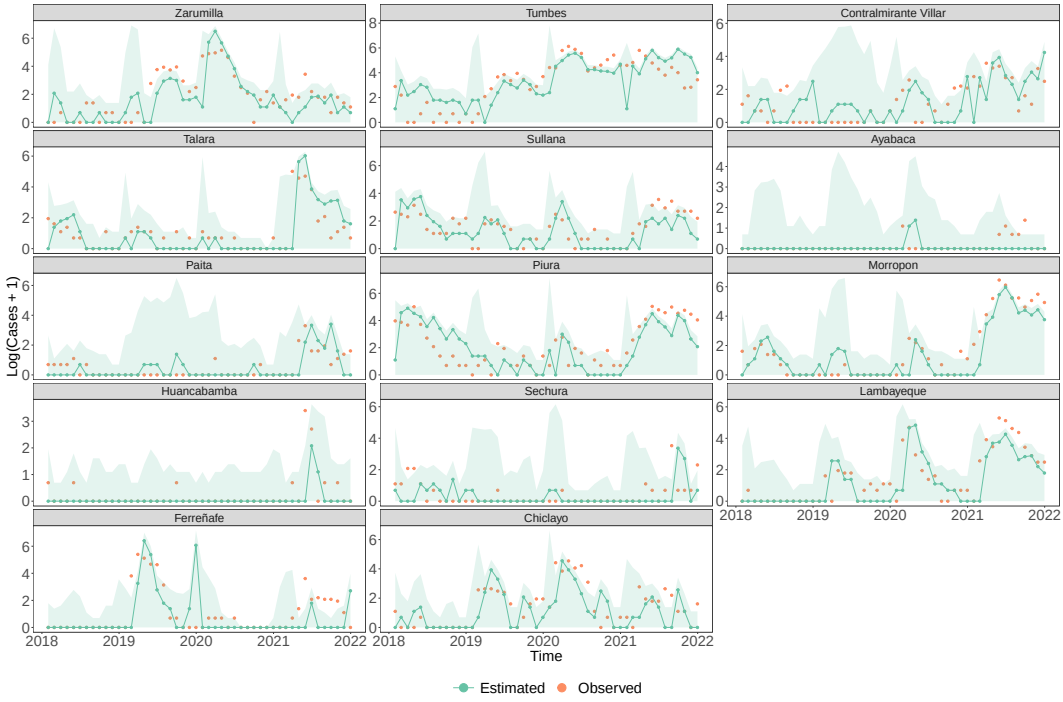


Figure SI C.30: **Bayesian climate model's predictive performance across the testing period (2018–2021).** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green) from one-month-ahead forecasts. The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics.

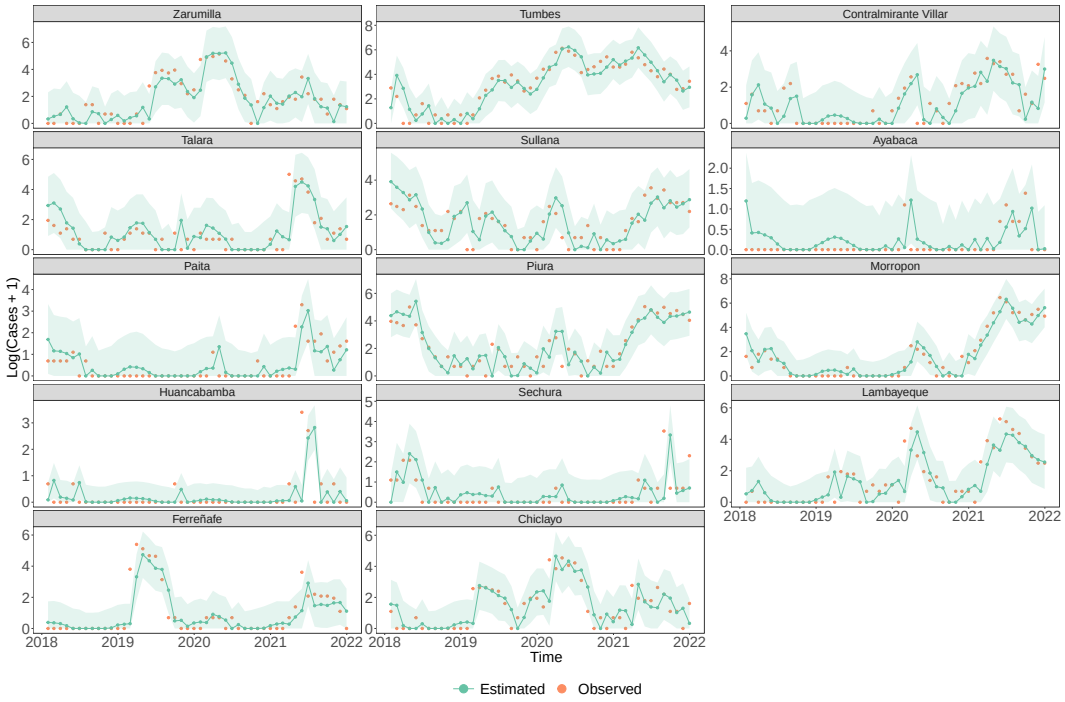


Figure SI C.31: **SARIMA models' predictive performance across the testing period (2018–2021).** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green) from one-month-ahead forecasts. The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics.



Figure SI C.32: **TCN models' predictive performance across the testing period (2018–2021).** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green) from one-month-ahead forecasts. The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics.

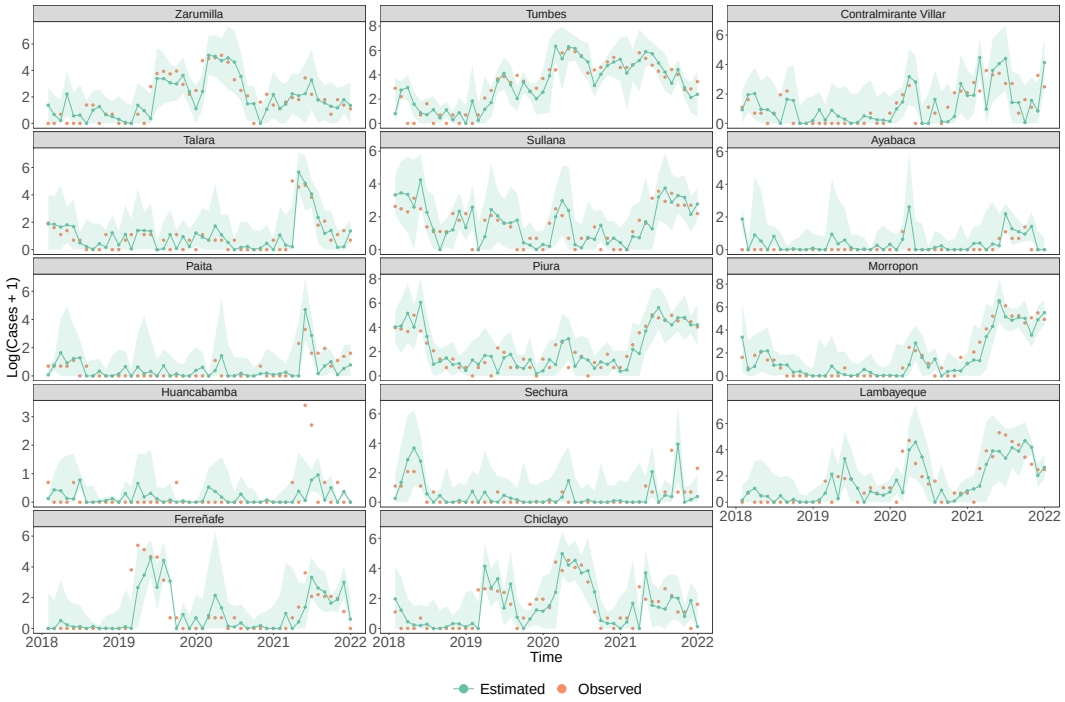


Figure SI C.33: **TimeGPT model's predictive performance across the testing period (2018–2021).** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green) from one-month-ahead forecasts. The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics.



Figure SI C.34: **Predictive performance of TimeGPT model without covariates across the testing period (2018–2021).** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green) from one-month-ahead forecasts. The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics.



Figure SI C.35: **Baseline model's predictive performance across the testing period (2018–2021).** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green) from one-month-ahead forecasts. The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics.

Testing period – Ensemble models

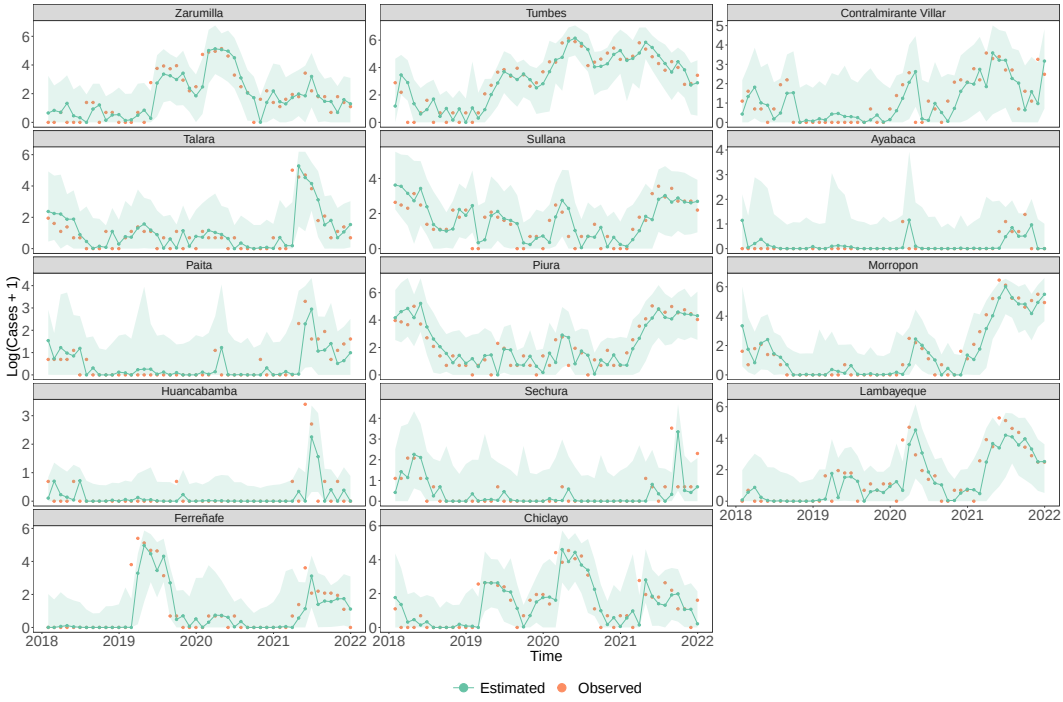
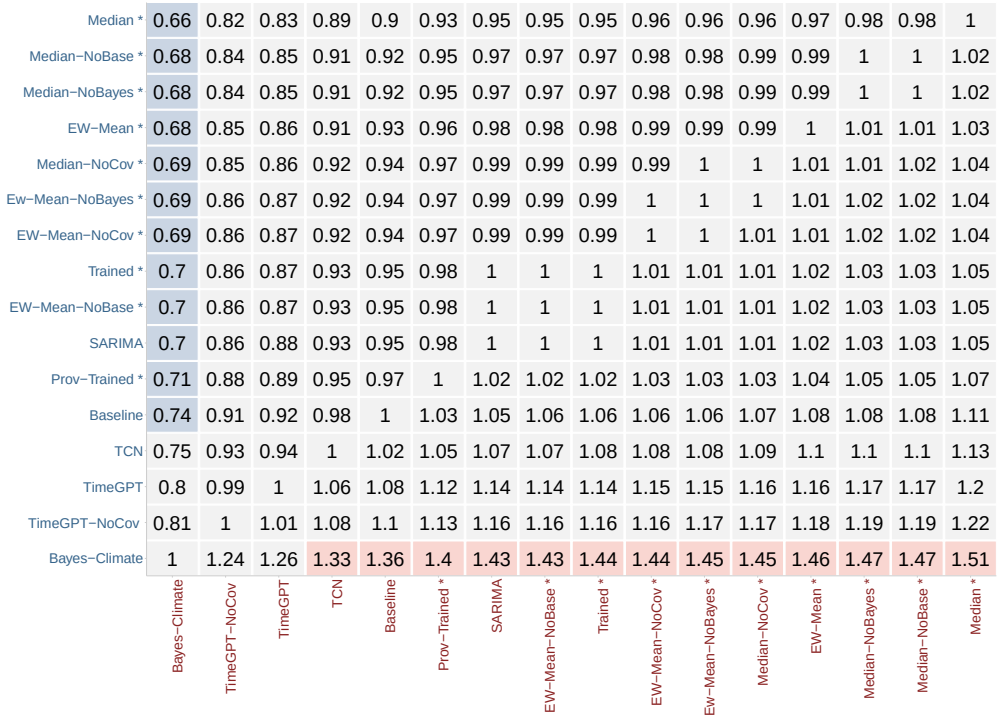


Figure SI C.36: **Median ensemble model's predictive performance across the testing period (2018–2021).** For each province, the observed cases (on a logarithmic scale) is shown (orange) alongside the corresponding posterior median estimates (green) and the 95% PIs (shaded green) from one-month-ahead forecasts. The provinces (facets) are sorted in descending order by latitude (from north to south), and y-axes are province-dependent due to differences in each province's average size of epidemics.

Testing period – WIS

Table SI C.14: **Summary statistics of WIS for log(Cases + 1) in the testing period of 2018 to 2021.** We summarise the weighted interval score (WIS) for each forecasting model by its median, lower quartile (Q1), and upper quartile (Q3). Models are arranged here by median WIS, where lower values of the WIS indicate superior forecasting performance. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk.

	Model	Median WIS	Q1	Q3
1	Median-NoBayes *	0.18	0.06	0.43
2	Median-NoCov *	0.18	0.06	0.44
3	Median *	0.18	0.06	0.42
4	TCN	0.18	0.05	0.54
5	Median-NoBase *	0.19	0.06	0.43
6	EW-Mean-NoCov *	0.19	0.08	0.43
7	Prov-Trained *	0.19	0.06	0.49
8	Ew-Mean-NoBayes *	0.20	0.08	0.43
9	TimeGPT-NoCov	0.20	0.07	0.53
10	SARIMA	0.20	0.07	0.41
11	Trained *	0.21	0.07	0.44
12	EW-Mean-NoBase *	0.21	0.07	0.45
13	EW-Mean *	0.22	0.08	0.43
14	TimeGPT	0.22	0.06	0.54
15	Baseline	0.22	0.06	0.51
16	Bayes-Climate	0.27	0.04	0.72



	Bayes−Climate	TimeGPT−NoCov	TimeGPT	TCN	Baseline	Prov−Trained *	SARIMA	EW−Mean−NoBase *	Trained *	EW−Mean−NoCov *	Ew−Mean−NoBayes *	Median−NoCov *	EW−Mean *	Median−NoBayes *	Median−NoBase *	Median *
Median *	0.66	0.82	0.83	0.89	0.9	0.93	0.95	0.95	0.95	0.96	0.96	0.96	0.97	0.98	0.98	1
Median−NoBase *	0.68	0.84	0.85	0.91	0.92	0.95	0.97	0.97	0.97	0.98	0.98	0.99	0.99	1	1	1.02
Median−NoBayes *	0.68	0.84	0.85	0.91	0.92	0.95	0.97	0.97	0.97	0.98	0.98	0.99	0.99	1	1	1.02
EW−Mean *	0.68	0.85	0.86	0.91	0.93	0.96	0.98	0.98	0.98	0.99	0.99	0.99	1	1.01	1.01	1.03
Median−NoCov *	0.69	0.85	0.86	0.92	0.94	0.97	0.99	0.99	0.99	0.99	1	1	1.01	1.01	1.02	1.04
Ew−Mean−NoBayes *	0.69	0.86	0.87	0.92	0.94	0.97	0.99	0.99	0.99	1	1	1	1.01	1.02	1.02	1.04
EW−Mean−NoCov *	0.69	0.86	0.87	0.92	0.94	0.97	0.99	0.99	0.99	1	1	1.01	1.01	1.02	1.02	1.04
Trained *	0.7	0.86	0.87	0.93	0.95	0.98	1	1	1	1.01	1.01	1.01	1.02	1.03	1.03	1.05
EW−Mean−NoBase *	0.7	0.86	0.87	0.93	0.95	0.98	1	1	1	1.01	1.01	1.01	1.02	1.03	1.03	1.05
SARIMA	0.7	0.86	0.88	0.93	0.95	0.98	1	1	1	1.01	1.01	1.01	1.02	1.03	1.03	1.05
Prov−Trained *	0.71	0.88	0.89	0.95	0.97	1	1.02	1.02	1.02	1.03	1.03	1.03	1.04	1.05	1.05	1.07
Baseline	0.74	0.91	0.92	0.98	1	1.03	1.05	1.06	1.06	1.06	1.06	1.07	1.08	1.08	1.08	1.11
TCN	0.75	0.93	0.94	1	1.02	1.05	1.07	1.07	1.08	1.08	1.08	1.09	1.1	1.1	1.1	1.13
TimeGPT	0.8	0.99	1	1.06	1.08	1.12	1.14	1.14	1.14	1.15	1.15	1.16	1.16	1.17	1.17	1.2
TimeGPT−NoCov	0.81	1	1.01	1.08	1.1	1.13	1.16	1.16	1.16	1.16	1.17	1.17	1.18	1.19	1.19	1.22
Bayes−Climate	1	1.24	1.26	1.33	1.36	1.4	1.43	1.43	1.44	1.44	1.45	1.45	1.46	1.47	1.47	1.51

Figure SI C.37: **WIS ratio between pairwise model combinations.** A visualisation of the pairwise ratio of Weighted Interval Score (WIS) values for forecasting models across 2018 to 2021. The focal model of interest is shown on the y-axis and the comparison models are shown on the x-axis. Then, the cell entries denote the ratio between the mean score of the focal model (numerator) and the mean score of the comparison model (denominator). Blue and red indicate superiority and inferiority of a model (as measured by the mean WIS) respectively. The p-value for each pairwise comparison is provided in Figure SI C.38, whilst the model abbreviations are described in Table SI C.2.

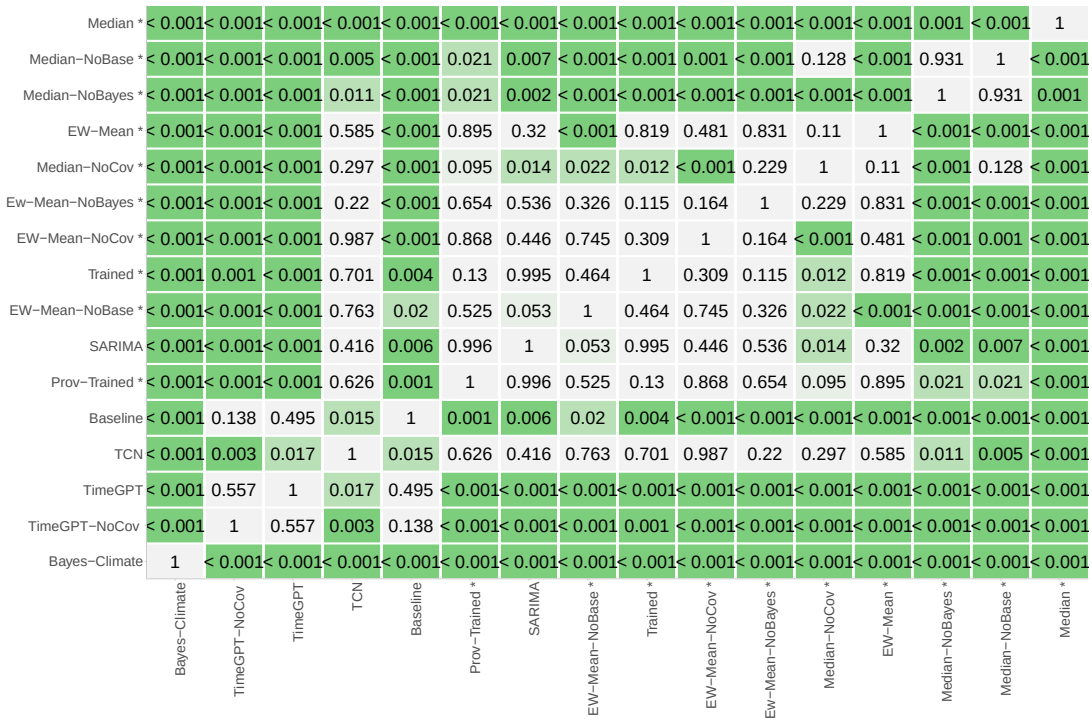


Figure SI C.38: **P-values for WIS ratio comparisons between pairwise model combinations.** P-values for the visualisation (Figure SI C.37) of the pairwise ratio of Weighted Interval Score (WIS) values for forecasting models across 2018 to 2021. The focal model of interest is shown on the y-axis and the comparison models are shown on the x-axis. Then, the cell entries denote the p-value for the ratio between the mean score of the focal model (numerator) and the mean score of the comparison model (denominator). The green of cell entries denotes statistically significance of the pairwise comparison as per a permutation test. The model abbreviations are described in Table SI C.2.

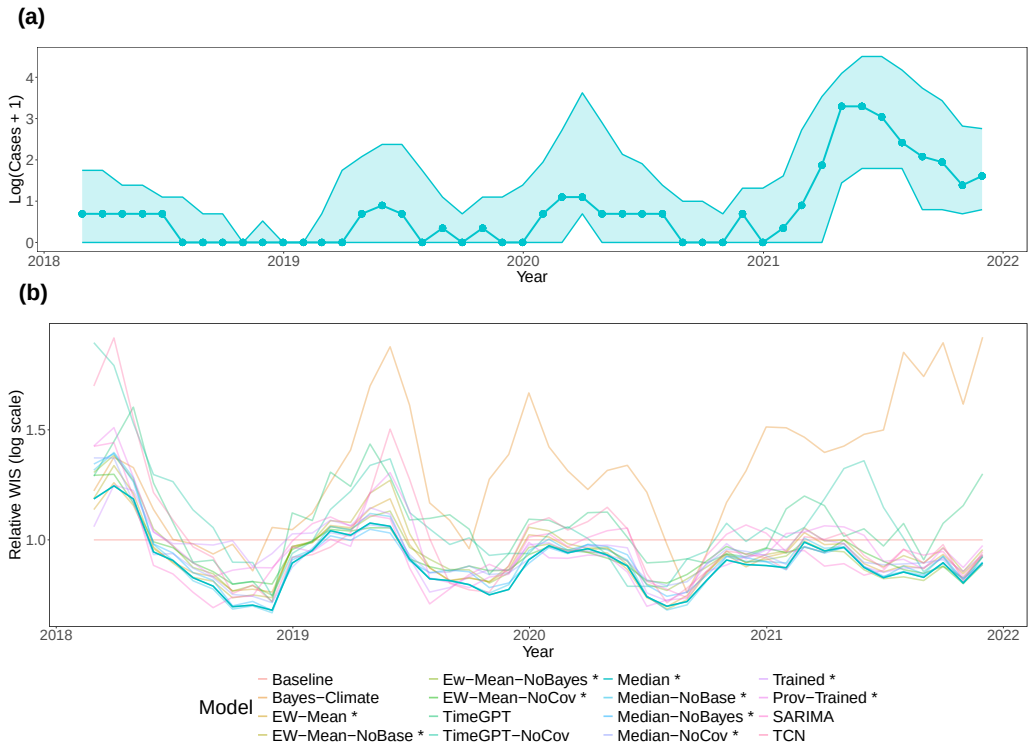


Figure SI C.39: **Relative WIS over time across 2018 to 2021.** (a) Across the fourteen provinces, the median, lower quartile (Q1), and upper quartile (Q3) are visualised for the observed log(Cases + 1) in centred, rolling windows of three months. (b) Relative Weighted Interval Score (WIS for log(Cases + 1)) is shown for all forecasting models in centred, rolling windows of three months, with the median model's values highlighted. Relative WIS is an individual forecasting models' WIS divided by the baseline forecasting's WIS, where values below one indicate greater forecasting skill relative to the baseline forecasting model. Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk.

Table SI C.15: **Correlations between relative WIS and median log(Cases + 1) in the testing period of 2018 to 2021.** The Pearson's correlation is reported, alongside exact 95% Confidence Intervals (CIs) calculated using the Clopper-Pearson Exact method [482]. Relative WIS is an individual forecasting models' WIS divided by the baseline forecasting's WIS, where lower values indicate greater forecasting skill relative to the baseline forecasting model. Higher positive correlation would indicate higher relative WIS (i.e. lower forecasting skill) for greater levels of log(Cases + 1). Details of the model abbreviations are provided in Table SI C.2. All ensemble models are highlighted with an asterisk.

	Model	Correlation	95% CI
1	Bayes-Climate	0.55	(0.31, 0.72)
2	EW-Mean *	0.07	(-0.22, 0.36)
3	EW-Mean-NoBase *	0.12	(-0.17, 0.40)
4	Ew-Mean-NoBayes *	-0.05	(-0.34, 0.24)
5	EW-Mean-NoCov *	0.02	(-0.27, 0.31)
6	TimeGPT	0.11	(-0.18, 0.39)
7	TimeGPT-NoCov	0.18	(-0.11, 0.45)
8	Median *	0.13	(-0.17, 0.41)
9	Median-NoBase *	0.12	(-0.18, 0.39)
10	Median-NoBayes *	0.08	(-0.21, 0.36)
11	Median-NoCov *	0.09	(-0.20, 0.37)
12	Trained *	-0.09	(-0.37, 0.21)
13	Prov-Trained *	0.17	(-0.13, 0.44)
14	SARIMA	-0.00	(-0.29, 0.29)
15	TCN	0.07	(-0.22, 0.35)

Testing period – Coverage

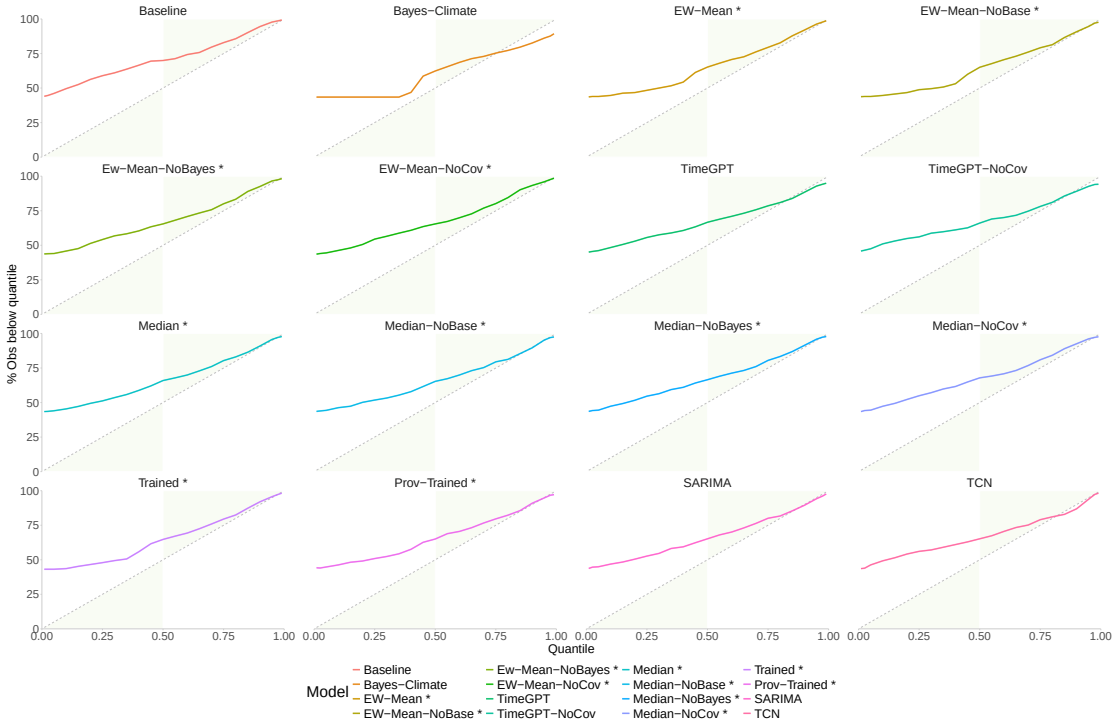


Figure SI C.40: **Quantile coverage for quantile-based forecasting models across 2018 to 2021.** For each forecasting model, the quantile coverage is plotted where the y-axis denotes the percentage of observed cases (on a logarithmic scale) that lie below the corresponding model-based predictive quantile level shown on the x-axis. Quantile coverage can be interpreted as a measure of probabilistic calibration i.e. consistency of a model's probabilities with those actually observed. The green shaded polygons denote areas where a model's predictive quantiles are conservative and cover more than needed. Details of the model abbreviations are provided in Table SI C.2.

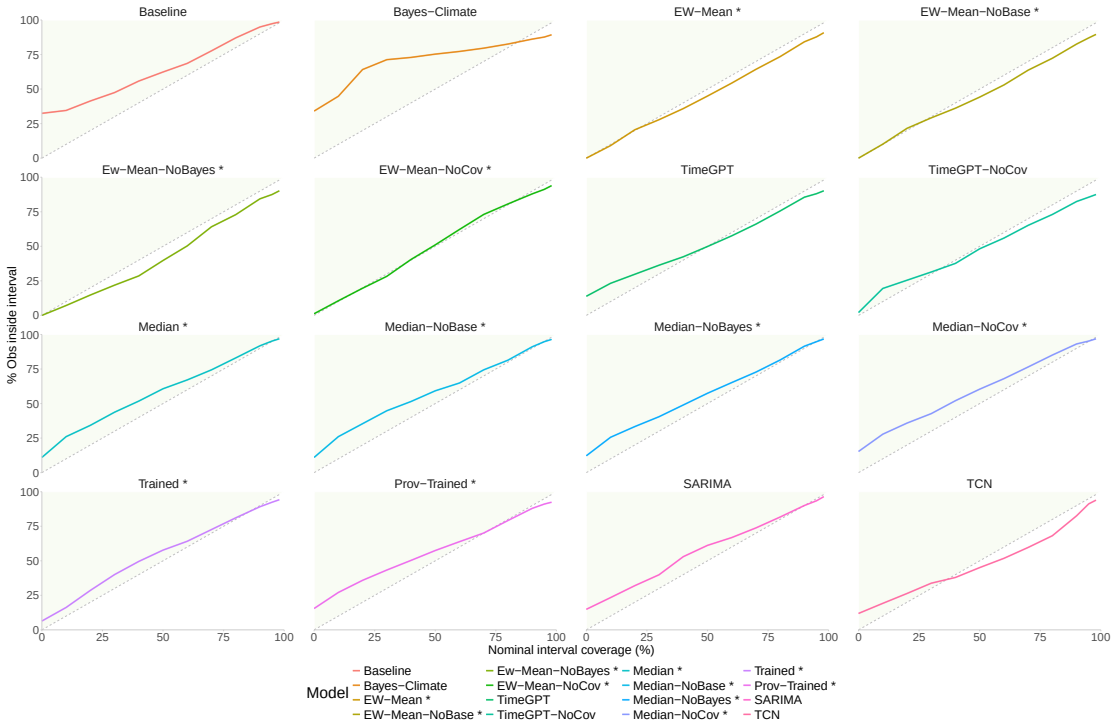


Figure SI C.41: **Interval coverage for quantile-based forecasting models across 2018 to 2021.** For each forecasting model, the interval coverage is plotted where the y-axis denotes the percentage of observed DIR values that lie corresponding model-based predictive quantile level shown on the x-axis. Interval coverage can be interpreted as a measure of probabilistic calibration i.e. consistency of a model's probabilities with those actually observed. The green shaded polygons denote areas of conservative behaviour i.e. empirical coverage outweighs the nominal coverage required. Interval coverage alone is not a measure of forecasting skill, as it only focuses on probabilistic calibration, and thus, we also report values of a proper scoring rule, the weighted interval score (WIS) Details of the model abbreviations are provided in Table SI C.2.

C.9.4 Additional figures

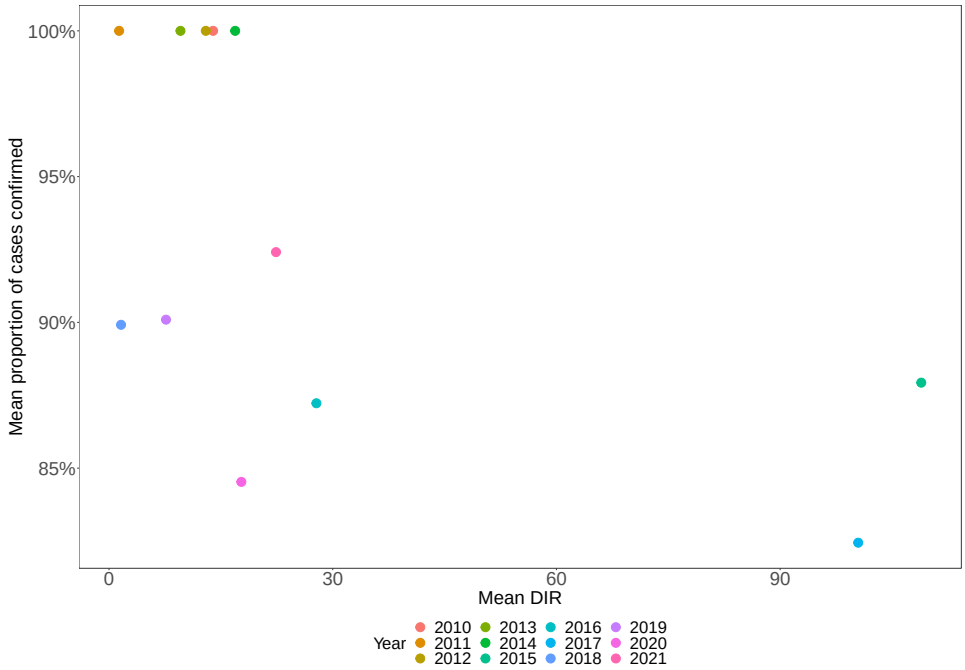


Figure SI C.42: **Relationship between mean proportion of cases confirmed and mean DIR.** Annual mean average of the monthly proportion of cases confirmed (y-axis) are plotted against the corresponding annual mean Dengue Incidence Rate (DIR) (y-axis). All three years with mean DIR ≥ 27.8 per 100,000 had average proportions of cases confirmed less than 88%.

Appendix D

Appendix — The time- and space-varying roles of human mobility in shaping urban dengue outbreaks

D.0.1 Notation and abbreviations

Table SI D.1: **Model abbreviations and description in Chapter 5**

Abbreviation	Definition
Bayesian	A Bayesian hierarchical model for dengue cases with mobility effects and neighbourhood-specific autoregressive processes for lagged cases.
No-Mob-Bayesian	A Bayesian hierarchical model for dengue cases with neighbourhood-specific autoregressive processes and no mobility effects for lagged cases.
DL	A foundational time series model (TimeGPT[401]) with mobility effects and autoregressive processes for lagged cases.
No-Mob-DL	A foundational time series model (TimeGPT[401]) with autoregressive processes for lagged cases.
Baseline	A baseline forecasting model for comparison and model validation (see [172,173]) .

Table SI D.2: **List of acronyms and descriptions in in Chapter 5.**

Acronym	Description
AR(1)	Autoregressive process of order one
CRPS	Continuous Ranked Probability Score
DENV	Dengue Virus
DIR	Dengue Incidence Rate
DL	Deep Learning
ELPD-LOO	Expected Log (pointwise) Predictive Density using Leave-One-Out (LOO) cross-validation
GAM	Generalised Additive Model
GCV score	Generalised Cross-Validation score
IBGE	Brazilian Institute of Geography and Statistics
LOO	Leave-One-Out cross-validation
MCMC	Markov Chain Monte Carlo
ML	Machine Learning
NegBin	Negative Binomial Distribution
$r_i(t)$	Ratio of mobility-generated contributions to a neighbourhood versus a neighbourhood's autoregressive process
$r\text{SHAP}_i(t)$	Our relative SHAP value for the estimated importance of human mobility effects
rWIS	Relative Weighted Interval Score
SHAP	SHapley Additive exPlanations (used for model interpretability)
SIMDA	Sistema de Monitoramento Diário de Agravos (Daily Disease Occurrence Monitoring System)
WAIC	Widely Applicable Information Criterion
WIS	Weighted Interval Score

D.1 Alternative candidate statistical models

We now consider alternative Bayesian hierarchical approaches to modelling y_{it}, the reported neighbourhood-level (i) weekly (t) case counts of dengue.

D.1.1 Statistical model I: Space-varying mobility effects

Our first framework included a mobility effect for each focal neighbourhood i, and weighted the past DIR per 100,000 of all neighbourhoods ($\eta_{j,t-1}$), including the focal neighbourhood itself, by the proportion of total flows originating from neighbourhood j to the focal neighbourhood i:

$$
\begin{aligned}
y_{it} &\sim NegBin(\mu_{it}, \phi), \\
\log(\mu_{it}) &= \log(P_{it}) + \log(\eta_{it}), \\
\log(\eta_{it}) &= \alpha + \left(\beta_{\text{Mobility}[i]} \cdot \sum_{j \in \mathcal{N}(i) \cup i} \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i) \cup i} \mathrm{f}_{t-1}^{(h,i)}} \cdot \eta_{j,t-1} \right) + \beta_{\text{Year}[t]} + \beta_{\text{Nbr}[i]}, \\
\beta_{\text{Mobility}[i]} &\sim N(\beta_{\text{Mobility}}, \sigma_{\text{Mobility}}), \\
\beta_{\text{Mobility}} &\sim N(0, 1), \\
\sigma_{\text{Mobility}} &\sim \text{Half-Normal}(0, 1), \\
\alpha &\sim N(0, 1), \\
\beta_{\text{Year}} &\sim N(0, 1), \\
\beta_{\text{Nbr}} &\sim N(0, 1), \\
\phi^{-1} &\sim \text{Half-Normal}(0, 1),
\end{aligned}
\tag{D.1.1}
$$

where μ_{it} denotes the mean parameter of the Negative Binomial distribution and ϕ is the reciprocal overdispersion parameter. We used a population offset $\log(P_{it})$ such that η_{it} represented the DIR per 100,000 population.

The DIR was then modelled with i) an intercept α, ii) a neighbourhood-specific mobility fixed effect controlled by $\beta_{\text{Mobility}[i]}$, the lagged DIR of all neighbourhoods ($\mathcal{N}(i) \cup i$ including the focal neighbourhood i), weighted by the inter- and intra-flows, iii) a year-to-year random effect (i.e. β_{Year} is a varying intercept for each year), and iv) a neighbourhood-specific random effect (i.e. β_{Nbr} is a varying intercept for each neighbourhood). So, $\beta_{\text{Mobility}[i]}$ is the neighbourhood-specific effect of past DIR across all neighbourhoods, weighted by mobility flows. Then, the common mean and standard deviation of these neighbourhood-specific effects are hyperparameters (β_{Mobility} and σ_{Mobility} respectively) which are given hyperprior distributions. All Normal prior distributions are presented in mean-standard deviation representation.

D.1.2 Statistical model II: Space-varying mobility with autoregressive processes

Our second model differs to that specified in eq. (D.1.1), as it allowed us to isolate the roles played by the autoregressive process of outbreaks within neighbourhoods and the mobility flows of (reported) infected individuals from other

neighbourhoods.

$$
\begin{aligned}
y_{it} &\sim NegBin(\mu_{it}, \phi), \\
\log(\mu_{it}) &= \log(P_{it}) + \log(\eta_{it}), \\
\log(\eta_{it}) &= \alpha + \left(\beta_{\text{Mobility-Nbr[i]}} \cdot \sum_{j \in \mathcal{N}(i)} \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}} \cdot \eta_{j,t-1} \right) + (\beta_{\text{AR[i]}} \cdot \eta_{i,t-1}) + \beta_{\text{Year}[t]} + \beta_{\text{Nbr}[i]}, \\
\beta_{\text{Mobility-Nbr}[i]} &\sim N(\beta_{\text{Mobility-Nbr}}, \sigma_{\text{Mobility-Nbr}}), \\
\beta_{\text{Mobility-Nbr}} &\sim N(0, 1), \\
\sigma_{\text{Mobility-Nbr}} &\sim \text{Half-Normal}(0, 1), \\
\beta_{\text{AR}[i]} &\sim N(\beta_{\text{AR}}, \sigma_{\text{AR}}), \\
\beta_{\text{AR}} &\sim N(0, 1), \\
\sigma_{\text{AR}} &\sim \text{Half-Normal}(0, 1), \\
\alpha &\sim N(0, 1), \\
\beta_{\text{Year}} &\sim N(0, 1), \\
\beta_{\text{Nbr}} &\sim N(0, 1), \\
\phi^{-1} &\sim \text{Half-Normal}(0, 1)
\end{aligned}
\tag{D.1.2}
$$

where μ_{it}, ϕ, P_{it}, η_{it}, α, β_{Year}, and β_{Nbr} are all defined as above. $\beta_{\text{Mobility-Nbr}[i]}$ is a neighbourhood-specific parameter controlling the effect of the weighted and lagged DIR of other neighbourhoods (excluding the focal neighbourhood), weighted by the inter-neighbourhood mobility flows. $\beta_{\text{AR[i]}}$ is the neighbourhood-specific parameter controlling a (potentially non-stationary) autoregressive process of order one, AR(1), for each neighbourhood i. As above, a hierarchical structure was imposed such that the mobility effect was modelled with partial pooling (common mean of $\beta_{\text{Mobility-Nbr}}$ and standard deviation of $\sigma_{\text{Mobility-Nbr}}$), and an identical procedure was performed for the autoregressive process (common mean of β_{AR} and standard deviation of σ_{AR}).

D.1.3 Statistical model III: Time-varying mobility with neighbourhood autoregressive effects

Let y_{it} be the reported neighbourhood-level (i) weekly (t) case counts of dengue which was modelled with a Negative Binomial (NegBin) likelihood as follows:

$$
\begin{aligned}
y_{it} &\sim NegBin(\mu_{it}, \phi), \\
\log(\mu_{it}) &= \log(P_{it}) + \log(\eta_{it}), \\
\log(\eta_{it}) &= \alpha + \left(\beta_{\text{Mobility-Nbr}[t]} \cdot \sum_{j \in \mathcal{N}(i)} \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}} \cdot \eta_{j,t-1} \right) + \beta_{\text{AR}[i]} \cdot \eta_{i,t-1} + \beta_{\text{Year}[t]} + \beta_{\text{Nbr}[i]}, \\
\beta_{\text{Mobility-Nbr}[1]} &\sim N(0, 1), \\
\beta_{\text{Mobility-Nbr}[t]} &\sim N(\beta_{\text{Mobility-Nbr}[t-1]}, \sigma_{\text{Mobility-Nbr}}), t \geq 2, \\
\sigma_{\text{Mobility-Nbr}} &\sim \text{Half-Normal}(0, 1), \\
\beta_{\text{AR}[i]} &\sim N(\beta_{\text{AR}}, \sigma_{\text{AR}}), \\
\beta_{\text{AR}} &\sim N(0, 1), \\
\sigma_{\text{AR}} &\sim \text{Half-Normal}(0, 1), \\
\alpha &\sim N(0, 1), \\
\boldsymbol{\beta}_{\text{Year}} &\sim N(0, 1), \\
\boldsymbol{\beta}_{\text{Nbr}} &\sim N(0, 1), \\
\phi^{-1} &\sim \text{Half-Normal}(0, 1),
\end{aligned}
\tag{D.1.3}
$$

where $\mu_{it}, \phi, P_{it}, \eta_{it}, \alpha, \beta_{\text{Year}}, \beta_{\text{Nbr}}, \beta_{\text{AR}[i]}, \beta_{\text{AR}}$ and σ_{AR} are all defined as above. Here, we modelled the effect of mobility to be spatially constant yet temporally evolving by allowing $\beta_{\text{Mobility-Nbr}}$ to follow a Gaussian Random Walk with standard deviation $\sigma_{\text{Mobility-Nbr}}$. Each neighbourhood was again given its own autoregressive process and varying intercept.

D.2 Results

D.2.1 Plots of data

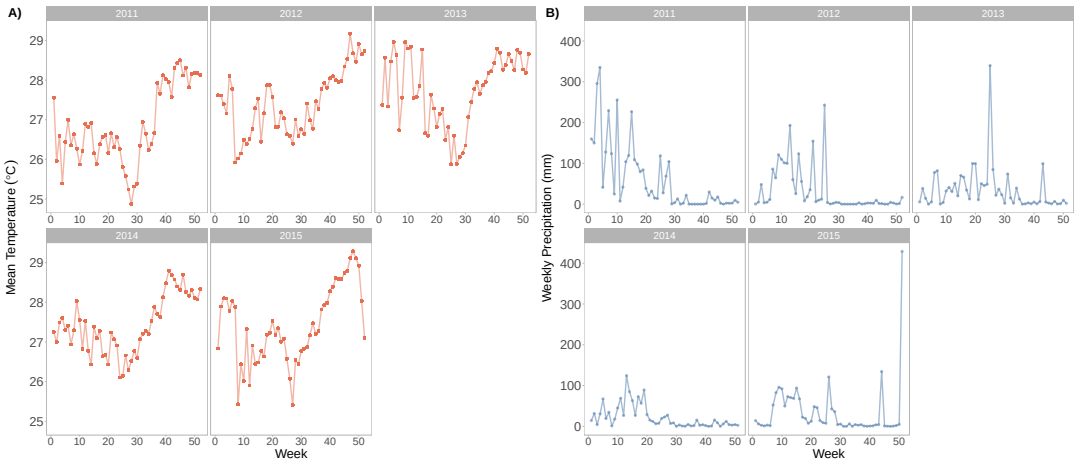


Figure SI D.1: **Weekly mean temperature and total precipitation in Fortaleza, 2011–2015: A)** Weekly mean temperature and **B)** weekly total precipitation for the city of Fortaleza, as per the ERA-5 dataset [282], where the variables were calculated as the average across the area within the Fortaleza city boundaries.

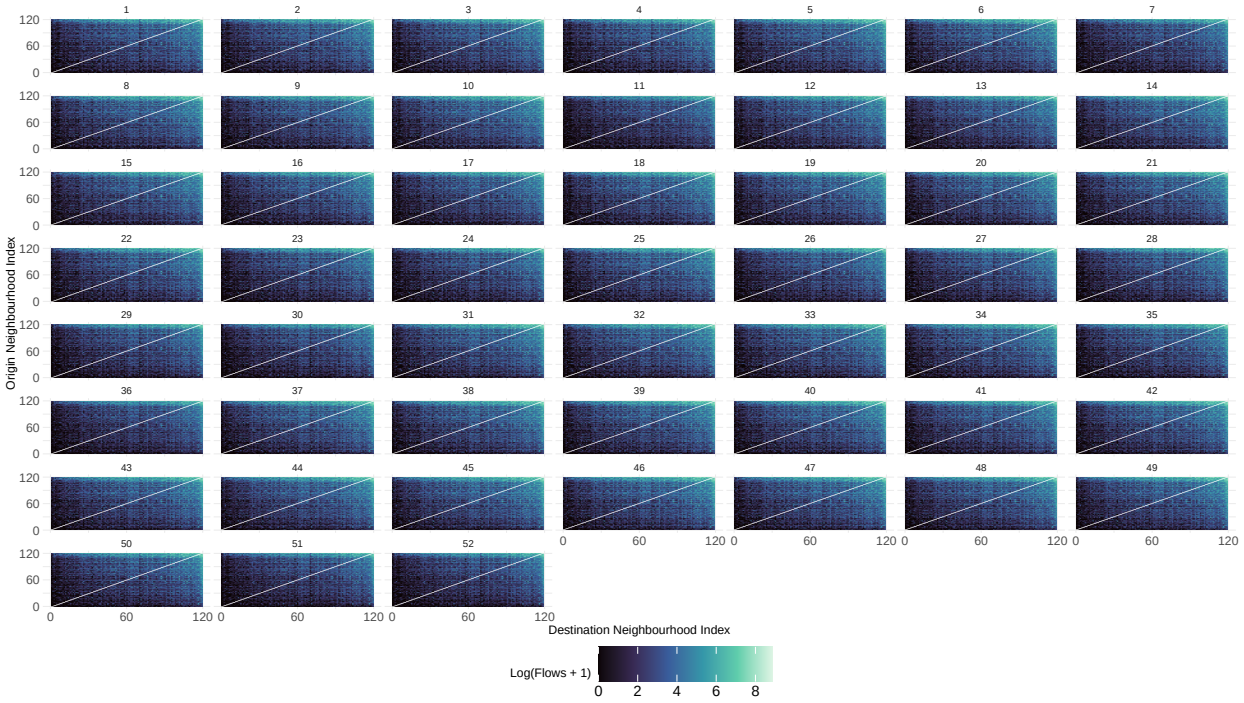


Figure SI D.2: **Weekly human movement in Fortaleza, Brazil:** The number of inflows and outflows between two neighbourhoods per week (facets) in 2015, where the neighbourhood index (for both axes) is consistent with other plots (for comparison purposes); ordered in ascending order by their total flows (inflows and outflows).

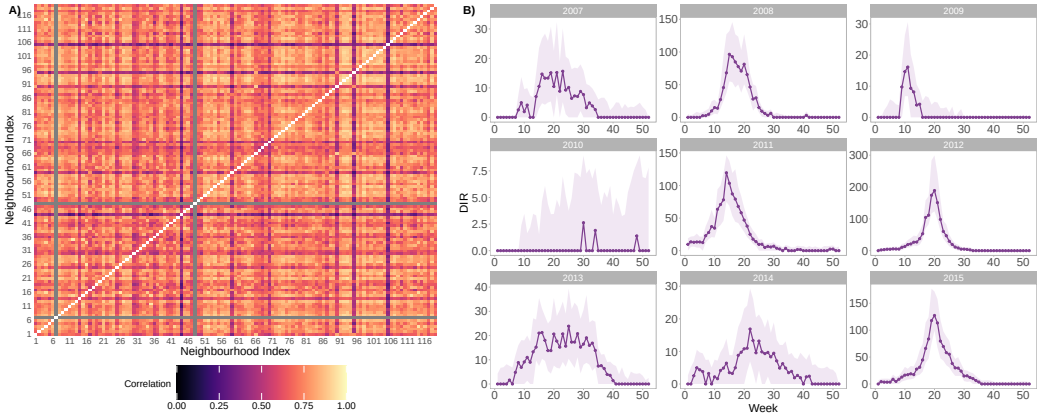


Figure SI D.3: **Epidemic synchrony and dengue cases from 2007–2015: A)** The Pearson's correlation coefficients[586] between the incident time series (using DIR for comparability) of each pair of the 119 neighbourhoods, where neighbourhoods are ordered in ascending order by their total flows (inflows and outflows). **B)** The median DIR (solid line) is visualised alongside upper and lower quartiles (ribbon) of DIR values for each time point across 2007–2015.

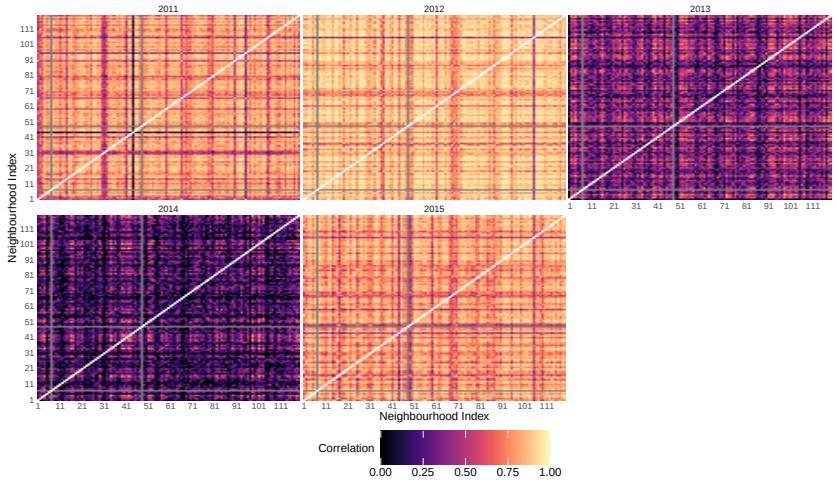


Figure SI D.4: **Within-year epidemic synchrony from 2007–2015:** The Pearson's correlation coefficient[586] between the incident time series (using DIR for comparability), per year, of each pair of the 119 neighbourhoods, where neighbourhoods are ordered in ascending order by their total flows (inflows and outflows). The epidemic synchrony for the full studied period is shown in Figure SI D.3A).

D.2.2 Time-varying relative importance of human mobility

The following section contains results which complement Section 5.3.2.

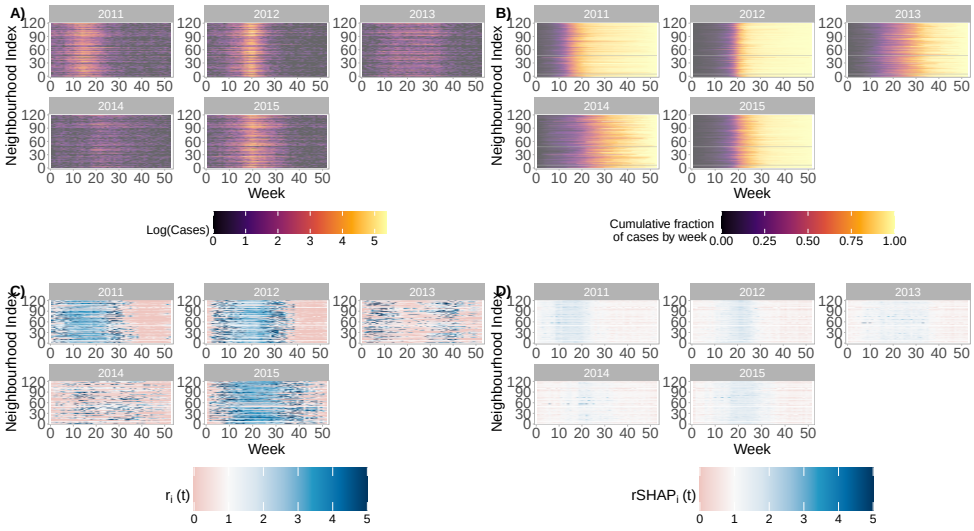


Figure SI D.5: **Time-varying relative importance of human mobility by year:** Across the testing period and for each of the 119 neighbourhoods, we visualise the estimated relative importance of human-movement-based importations of dengue cases, versus the neighbourhoods' own autoregressive/internal generation of new dengue cases. **A)** and **B)** visualise the observed cases (on a logarithmic scale) and cumulative fraction of annual cases observed for individual weeks across 2011-2015. Neighbourhoods are ordered in ascending order by their total annual number of flows (inflows and outflows). **C)** and **D)** show the estimated relative importance from our Bayesian hierarchical and deep learning models respectively, where values of $r_i(t)$ and $r\text{SHAP}_i(t)$ greater than one indicate greater relative importance of the case importations (see Materials and methods). Figure 5.3 C–D is the analogous continuous version of these estimates of $r_i(t)$ and $r\text{SHAP}_i(t)$.



Figure SI D.6: **Time-varying relative importance of human mobility by different models, alongside observed case data: A)** Median neighbourhood-level dengue incidence rate (DIR) on a logarithmic scale is visualised alongside **B)** $r_i(t)$ and $rSHAP_i(t)$, the model-based estimates for the time-varying relative importance of human mobility. Shaded intervals visualise the interquartile range.

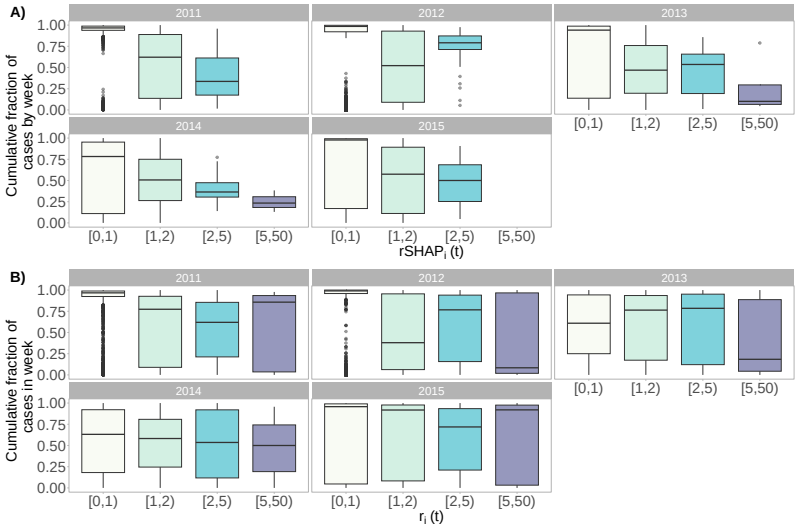


Figure SI D.7: **Time-varying importance of human mobility by epidemic stage in different years:** For each year (2007–2015), a summary of the cumulative fraction of annual cases observed by a given week is visualised against the **A)** corresponding deep learning model's estimates and **B)** the Bayesian hierarchical model's estimates for the relative importance of human mobility for each neighbourhood's epidemic dynamics. The box and whiskers capture the interquartile range.



Figure SI D.8: **Bayesian model's estimated time-varying importance of human mobility, by epidemic stage in different neighbourhoods:** For each neighbourhood (indexed 1 to 119, in ascending order by mobility flows), the cumulative fraction of annual cases observed by a given week is visualised against the corresponding Bayesian hierarchical model's estimates for the relative importance of human mobility for the neighbourhood's epidemic dynamics. The blue line represents the conditional mean and the shaded band represents the 95% confidence interval (calculated using bootstrap resampling) for each $r_i(t)$, calculated using loess (Local Polynomial Regression Fitting).



Figure SI D.9: **Deep learning model's estimated time-varying importance of human mobility, $r\text{SHAP}_i(t)$ by epidemic stage in different neighbourhoods:** For each neighbourhood (indexed 1 to 119, in ascending order by mobility flows), the cumulative fraction of annual cases observed by a given week is visualised against the corresponding Bayesian hierarchical model's estimates for the relative importance of human mobility for the neighbourhood's epidemic dynamics. The blue line represents the conditional mean and the shaded band represents the 95% confidence interval (calculated using bootstrap resampling) for each $r\text{SHAP}_i(t)$, calculated using loess (Local Polynomial Regression Fitting).

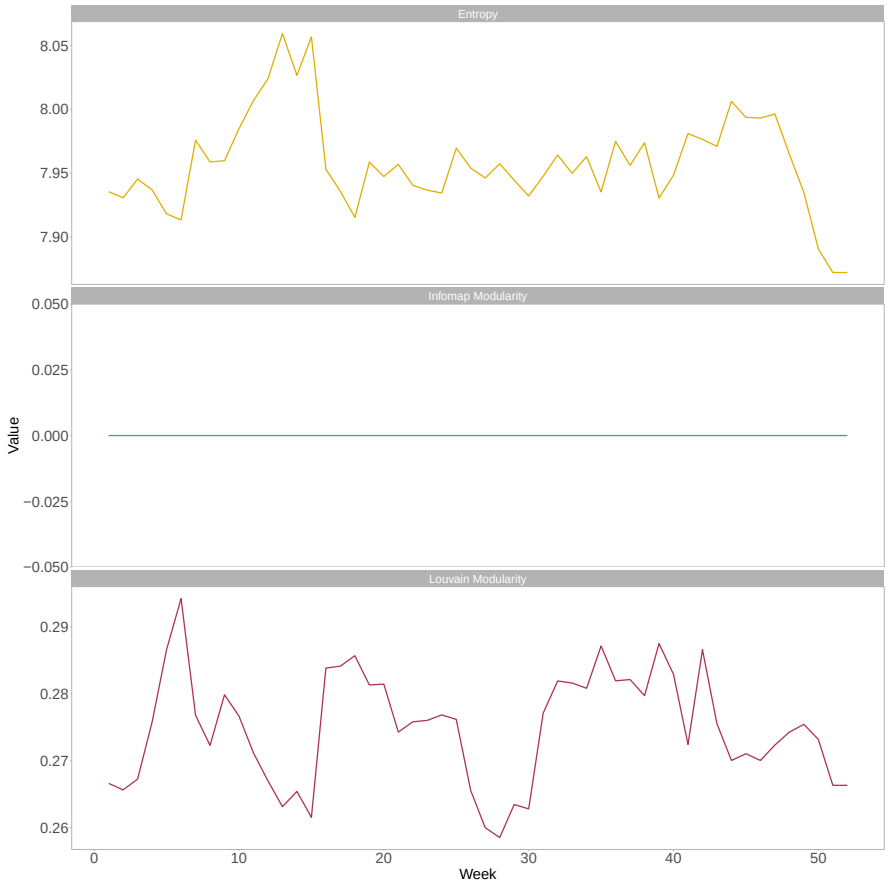


Figure SI D.10: **Mixing patterns in Fortaleza by different algorithms:** We quantified mobility clustering using entropy, Infomap modularity, and Louvain modularity, where higher values of entropy and lower values of modularity indicate greater mixing and less mobility clustering. The entropy-based results are constrained by an upper bound of $\log(119^2) = 9.56$, the Infomap modularity indicates no clusters identified, and Louvain modularity ranges from -1 to 1, with values close to zero indicating low clustering.

D.2.3 Space-varying human-mobility-driven roles of neighbourhoods

The following section contains results which complement Section 5.3.3.

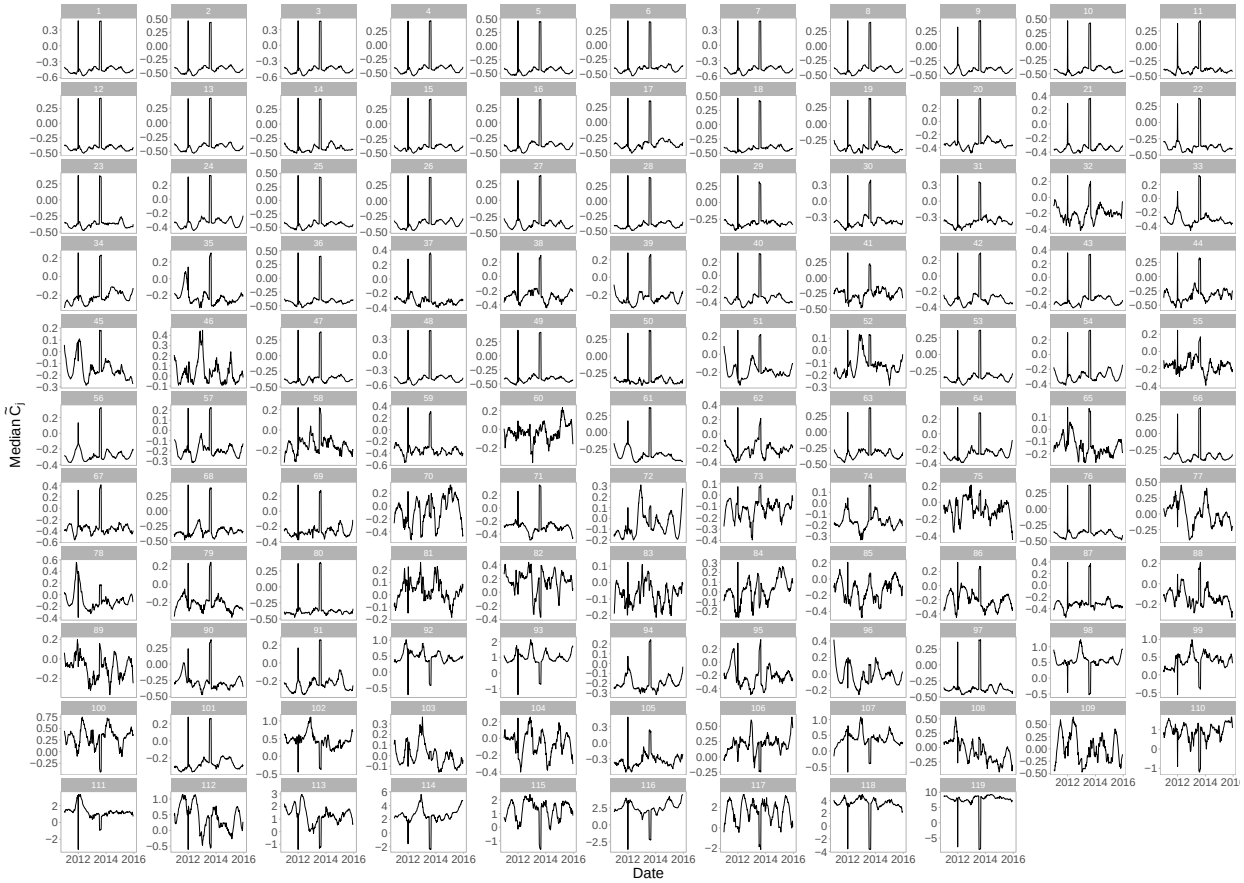


Figure SI D.11: **Estimated time-varying contributions of neighbourhoods:** For each neighbourhood j, $\widetilde{C_j}$, is shown where the date on the x-axis represents the maximum date for that 17-week trailing window. $\widetilde{C_j}$ is the Bayesian model's estimate for the standardised contribution of neighbourhoods j to the dengue epidemics of all other neighbourhoods. This can be compared across different rolling windows due to the standardisation. Note the different y-axis to preserve the different magnitudes of contributions from different neighbourhoods.

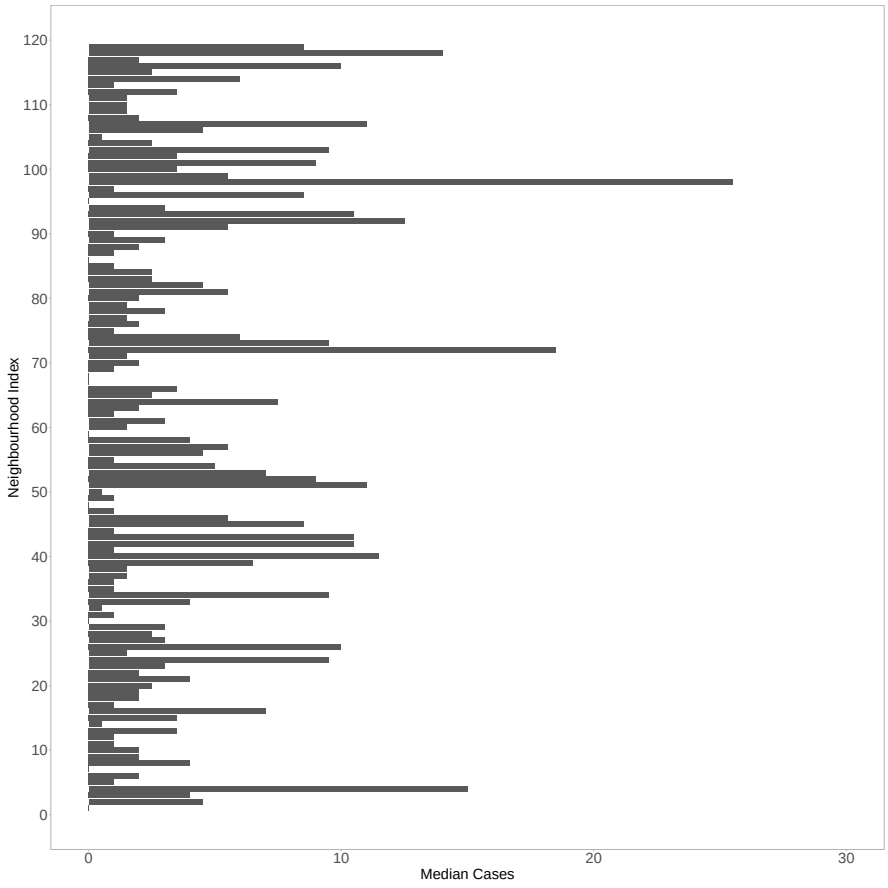


Figure SI D.12: **Median weekly cases by neighbourhood during 2013 season:** For each neighbourhood, we examined the reported incident cases during a subset (July–August 2013 inclusive) of the abnormal, low-incidence season of 2013, as we had estimated lower relative contributions of high-mobility neighbourhoods during this time period. Neighbourhoods are ordered in ascending order from lowest to highest total mobility flows.

Appendix E

Appendix — Reproduction numbers for vector-borne diseases

E.1 Renewal equations, generation times, and McKendrick–von Foerster equations

The epidemic renewal equation framework forms the backbone of many popular methods for inferring transmissibility throughout the course of an infectious disease outbreak [137,139,140,148,149]. Mathematically, the renewal equation framework (in continuous time) for directly transmitted human-to-human infectious diseases can be written as:

$$i(t) = R(t) \int_0^\infty i(t-\tau) w(\tau) d\tau \tag{E.1.1}$$

where the instantaneous reproduction number $R(t)$ is the expected number of secondary human infections generated per primary human infection if conditions were to remain the same as at time t, and $i(t-\tau)$ is the number of human infections that occurred at time $t-\tau$ which is weighted by the probability density of the generation time $w(\tau)$ (by convention, $w(\tau) = 0$ for $\tau < 0$ and $\int_0^\infty w(\tau) d\tau = 1$). The framework is most often implemented in discrete time to align with the temporal resolution of surveillance data.

Whilst the generation time distribution can be viewed as the distribution of times between primary human infection and secondary transmission, it is also useful to view the generation time distribution as the expected relative contribution to the force of infection by a human who was infected a specific number of time units ago (here, τ units ago) [145,606]. The latter interpretation is useful for our proposed frameworks. It is important to distinguish that the renewal equation, eq. (E.1.1), requires the intrinsic generation time distribution (i.e., the time between primary human infection and secondary transmission assuming constant transmission conditions), as opposed to other epidemiological intervals (such as transmission intervals, serial intervals, observed generation times, etc.).

The generation time is a key (and often overlooked [107,113]) modelling ingredient for any renewal equation framework – its distribution controls times between a primary human infection and secondary human infection. However, when using renewal equation frameworks, the generation time distribution is commonly assumed to be known/fixed (and not estimated), and is often approximated using the serial interval (the time from symptom onset of the primary case to symptom onset in the secondary case). The use of a time-homogeneous generation time distribution (fixed over calendar time), as in eq. (E.1.1), can often be justified (partially) using both statistical (changes to $R(t)$ and w are often not simultaneously identifiable from available surveillance data) and epidemiological (as an intrinsic generation time distribution is time-invariant in the

setting of the constant-strength changes which only impact the epidemic strength – $R(t)$) arguments. Nevertheless, the assumption of a fixed, known generation time distribution means that misspecification of the generation time distribution can be a large source of bias when estimating $R(t)$ and of the true branching process dynamics of an epidemic [113,134,142,145]. Given the popularity of $R(t)$ for informing public health policy and decision-making [118], the violation of the assumption of time-invariant generation times clearly has practical, real-world consequences.

Time-varying (also known as instantaneous or time-inhomogeneous) generation time distributions are less commonly used. These allow for the relative contribution of individuals with specific ages (i.e., times since infection) to vary over calendar time, and can be derived as the normalised version of the instantaneous kernel [145]:

$$g(t,\tau) := \frac{K(t,\tau)}{R(t)}$$

where $K(t,\tau)$ is the instantaneous kernel which describes the rate at which secondary human infections are generated at calendar time t by individuals who were infected τ time units ago, and $R(t) = \int_0^\infty K(t,\tau)d\tau$ is the instantaneous reproduction number which has the same interpretation as in eq. (6.1.1).

For vector-borne pathogens such as dengue, the assumption of a fixed generation time distribution over calendar time is highly unrealistic due to mosquito-viral traits (such as the extrinsic incubation period, the rate of mosquito blood feeding, and mosquito density) which have estimated potentially strong temperature dependencies [80,82,83]. Similarly, the length of time between primary and secondary human infections is likely influenced by the totals of susceptible and infectious vectors and host. So, in addition to the multi-stage nature of the transmission cycle, the ubiquitous renewal equation framework, eq. (E.1.1), is not directly applicable for vector-borne diseases.

Focusing on dengue, recent works (outside renewal equation frameworks) have leveraged empirical data to derive time-varying, temperature-dependent generation time distributions. For instance, the generation time distribution has been modelled as a convolution of temperature-dependent parametric distributions for each of the four stages of the dengue transmission cycle (i.e., from the primary human infection to subsequent human transmission). Standard Ross-Macdonald, temperature-dependent assumptions for mosquito-borne pathogens [121,325] were used to calculate a temperature-driven time-varying reproduction-number R_0 (and exponential growth rate r_t) from elements derived from the generation time (such as driven death rates, mosquito-to-human ratios, and the extrinsic incubation period, all of which were determined by temperature) [156]. Similarly, another framework uses time-varying, temperature-dependent generation time distributions and the same empirical data and constituent phases of the transmission cycle, albeit with alternative distributional forms (gamma distributions) for mathematical convenience [375]. The generation time distribution was determined in a piecewise constant, recursive form by assuming temperature varies in discrete steps (weekly here). Then, an instantaneous reproduction number $R(t)$ at time t was derived using the assumptions of [133], the case count data up to time t, and the derived time-varying, temperature-driven generation time distribution. Other mechanistic approaches have also regularly deployed extensions of the Ross-Macdonald framework, yielding basic reproduction numbers determined by temperature, as above and in [83], and basic reproduction numbers determined by temperature and vector carrying capacity per human [601]. Nevertheless, across each of the existing methods, there has been a strong dependence on temperature forcing. Furthermore, there has been a lack of focus on a clear mechanistic, mathematical account of how the derived instantaneous reproduction number R is determined by the calendar-time-varying properties of the human and vector sub-populations that form the stages of the transmission cycle of the vector-borne pathogen. Similarly, there is no clear understanding of what it means to assume a temporally evolving generation time distribution for each constituent phase and sub-population involved in the human-to-human transmission cycle.

E.2 Exposure-and-infection-age-structured human and infection-age-structured mosquito population

The following section presents additional details of derivations for the framework initially discussed in Section 6.2.1.

E.2.1 Modelling framework

Recall that we previously had the following system of age-structured populations at calendar times t, with E and M denoting the density of the infected (and here, also infectious) human and mosquito populations, respectively (with corresponding ages a_E and a_M):

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a_E} = 0, \tag{E.2.1}$$

$$\frac{\partial M}{\partial t} + \frac{\partial M}{\partial a_M} = -\mu M(t, a_M), \tag{E.2.2}$$

where $\mu > 0$ is the (constant) mortality rate in the (infected) mosquito population and we close the system with the following initial and boundary conditions:

$$E(t, 0) = \int_0^\infty b_{ME}(t, a_M) M(t, a_M) da_M, \tag{E.2.3}$$

$$E(0, a_E) = f_E(a_E), \tag{E.2.4}$$

$$M(t, 0) = \int_0^\infty b_{EM}(t, a_E) E(t, a_E) da_E, \tag{E.2.5}$$

$$M(0, a_M) = f_M(a_M). \tag{E.2.6}$$

where $f_E(a_E)$ and $f_M(a_M)$ denote the initial densities of infectious human and mosquito populations, and each $b_{AB}(t, a_A) \geq 0$ is a time- and age-dependent birth function that represents the rate at which new individuals in sub-population B are generated at calendar time t by an individual of age a_A in sub-population A. As a linear PDE with suitable initial and boundary conditions creates a Cauchy problem, we can now derive formulae for renewal equations, instantaneous reproduction numbers, and time-varying generation time distributions using the method of characteristics (which tries to recreate the solution surface of our systems).

E.2.2 Deriving a renewal equation and instantaneous reproduction number

Now, we provide additional details of the derivation for our characteristic equation, eq. (6.2.8). First, considering the characteristics for the mosquito population, we have

$$a_M = \begin{cases} t - t_0, & \text{if } a_M \leq t, \\ t + a_0, & \text{if } a_M > t. \end{cases} \tag{E.2.7}$$

where t_0 is the time at which an individual mosquito of age $0 < a_M < t$ is born, and a_0 is the initial age of an individual mosquito who has age $a_M > t$ at time t. For each of the two cases given by eq. (E.2.7), we can determine the solution:

$$M(t, a_M) = \begin{cases} M(t - a_M, 0) e^{-\mu a_M}, & \text{if } a_M \leq t, \\ f_M(a_M - t) e^{-\mu t}, & \text{if } a_M > t. \end{cases} \tag{E.2.8}$$

This means that we can rewrite eq. (E.2.3) as:

$$E(t,0) = \int_0^t b_{ME}(t,a_M)M(t-a_M,0)e^{-\mu a_M}da_M + \int_t^\infty f_M(a_M-t)e^{-\mu t}da_M, \tag{E.2.9}$$

where we are mostly concerned with the long-time behaviour of the system, i.e., when $t \gg a_M$ and $f_M(a_M-t)e^{-\mu t} \approx 0$. Then we can neglect the second part of eq. (E.2.9) and rewrite it as:

$$E(t,0) = \int_0^t b_{ME}(t,a_M)M(t-a_M,0)e^{-\mu a_M}da_M. \tag{E.2.10}$$

We finally substitute eq. (E.2.5) into eq. (E.2.10) to yield the renewal equation:

$$E(t,0) = \int_0^t \int_0^\infty b_{ME}(t,a_M)e^{-\mu a_M}b_{EM}(t-a_M,a_E)E(t-a_M,a_E)da_E da_M. \tag{E.2.11}$$

eq. (E.2.11) makes explicit the fact that mosquito mortality needs to be taken into account when using renewal equations to model vector-borne diseases.

In the same way as for the mosquito population, we can write the solution for the human population as follows:

$$E(t,a_E) = \begin{cases} E(t-a_E,0), & \text{if } a_E \le t, \\ f_E(a_E-t), & \text{if } a_E > t. \end{cases} \tag{E.2.12}$$

This means that, focusing on long-term behaviour of the system (i.e., if t is so large ($t \to \infty$) such that $t > a_E$, then $f_E(a_E-t) = 0$.), we can rewrite eq. (E.2.11) as

$$E(t,0) = \int_0^t \int_0^{t-a_M} b_{ME}(t,a_M)e^{-\mu a_M}b_{EM}(t-a_M,a_E)E(t-a_M-a_E,0)da_E da_M. \tag{E.2.13}$$

We now seek a long-term solution of the form: $E(t,a_E) = p(a_E)e^{\gamma t}$. This solution, defined instantaneously at time t using long-run behaviour, states that the age distribution of exposed individuals is altered by a factor which grows or decays with time depending on whether $\gamma > 0$ or $\gamma < 0$, respectively, if conditions affecting human infections were to remain the same. We first substitute this into eq. (E.2.1) to yield:

$$\gamma p(a_E)e^{\gamma t} + \frac{dp(a_E)}{da_E}e^{\gamma t} = 0, \tag{E.2.14}$$

which implies

$$p(a_E) = p(0)e^{-\gamma a_E}. \tag{E.2.15}$$

Now substituting the long-term solution and eq. (E.2.15) into eq. (E.2.11) yields:

$$p(0)e^{\gamma t} = \int_0^t \int_0^{t-a_M} b_{ME}(t,a_M)e^{-\mu a_M}b_{EM}(t-a_M,a_E)p(0)e^{\gamma(t-a_M-a_E)}da_E da_M, \tag{E.2.16}$$

which upon cancellation of $p(0)e^{\gamma t}$ from both sides gives:

$$1 = \int_0^t \int_0^{t-a_M} b_{ME}(t,a_M)e^{-\mu a_M}b_{EM}(t-a_M,a_E)e^{-\gamma(a_E+a_M)}da_E da_M := \phi(\gamma). \tag{E.2.17}$$

E.2.3 Relation to existing renewal equation frameworks

We provide here the additional details for how our renewal equation, eq. (6.2.13), can be related to existing renewal equation frameworks for directly transmitted infectious diseases (where a homogeneous delay distribution over calendar time is typically assumed).

Denoting $\tau = a_E + a_M$ as the overall delay between primary and secondary human infections, the relation (and compression) of the density for our generation time $g(t, a_M, a_E)$ to a 1-D marginal density $w(t, \tau)$ at time t), can be established by i) a Jacobian-based change of variable (to parameterise our integral equation at time t in terms of τ) and ii) $C - 1 = 1$ integration steps (where $C = 2$ is the number of model compartments) to marginalise out the first times-between-state variable (a_E).

Focusing on the integral

$$\int_0^t \int_0^{t-a_M} g(t, a_M, a_E) E(t - a_M - a_E, 0) da_E da_M$$

from eq. (6.2.13), we will now use a change of variables to convert $(a_E, a_M) \to (a_E, \tau)$. Using bijective functions $l(a_E, \tau) = a_M = \tau - a_E$, and $m(a_E, \tau) = a_E$, we have

$$\begin{aligned}
&\int_0^t \int_0^{t-a_M} g(t, a_M, a_E) E(t - a_M - a_E, 0) da_E da_M \\
&= \int_0^t \int_0^{\tau} g(t, l(a_E, \tau), m(a_E, \tau)) E(t - l(a_E, \tau) - m(a_E, \tau), 0) \left| \frac{\partial(a_E, a_M)}{\partial(a_E, \tau)} \right| da_E d\tau \\
&= \int_0^t \int_0^{\tau} g(t, \tau - a_E, a_E) E(t - (\tau - a_E) - a_E, 0) da_E d\tau \\
&= \int_0^t E(t - \tau, 0) \int_0^{\tau} g(t, \tau - a_E, a_E) da_E d\tau \\
&= \int_0^t E(t - \tau, 0) \left[\tilde{g}(t, \tau - a_E, a_E) \right]_0^{\tau} d\tau \\
&= \int_0^t E(t - \tau, 0) w(t, \tau) d\tau
\end{aligned} \tag{E.2.18}$$

where $\left| \frac{\partial(a_E, a_M)}{\partial(a_E, \tau)} \right| = 1$ denotes the Jacobian (i.e., the determinant of the Jacobian matrix), $\tilde{g}(t, \tau - a_E, a_E)$ is the antiderivative of $g(t, \tau - a_E, a_E)$ with respect to a_E, and $w(t, \tau)$ is the corresponding definite integral defined as $w(t, \tau) := [\tilde{g}(t, \tau - a_E, a_E)]_0^{\tau}$ (which is again evaluated with respect to a_E). Note that $g(t, l(a_E, \tau), m(a_E, \tau))$ was rewritten as $g(t, \tau - a_E, a_E)$.

Then, our renewal equation becomes

$$E(t, 0) = R(t) \int_0^t w(t, \tau) E(t - \tau, 0) d\tau. \tag{E.2.19}$$

where $w(t, \tau) := [\tilde{g}(t, \tau - a_E, a_E)]_0^{\tau} = \int_0^{\tau} g(t, \tau - a_E, a_E) da_E$.

If we additionally assume a homogeneous generation time distribution over calendar time, (i.e., $g(t, a_M, a_E) = g(a_M, a_E) \forall t$, we would have the familiar renewal equation framework:

$$E(t,0) = R(t) \int_0^t w(\tau) E(t-\tau, 0) d\tau. \tag{E.2.20}$$

Assuming a homogeneous generation time distribution over calendar time would mean that the ratio defined by

$$g(t, a_M, a_E) = \frac{K(t, a_M, a_E)}{R(t)} = \frac{b_{ME}(t, a_M) e^{-\mu a_M} b_{EM}(t - a_M, a_E)}{\int_0^t \int_0^{t-a_M} b_{ME}(t, a_M) e^{-\mu a_M} b_{EM}(t - a_M, a_E) da_E da_M}, \tag{E.2.21}$$

is fixed over calendar time (t). So, the typically implicit modelling assumption means that the relative contribution (to secondary human infections) of each times-since-compartment-entry pair (a_M, a_E) to new current human infections is constant throughout the course of an outbreak. In the current framework, such an assumption means that the expected contribution (to new human infections) of each *individual* pair of times-in-states ($K(t, a_M, a_E)$) scales linearly over calendar time t with the expected *overall* number of new infections per initially infected human ($R(t)$).

The assumption may be unrealistic for a vector-borne disease with a multi-stage transmission cycle such as dengue (e.g., higher temperatures may reduce the extrinsic incubation period[80] and/or changes in the host and vector population sizes may impact the distribution of times between generation events).

Nevertheless, eq. (E.2.20) formalises the relationship between our framework and the conventional familiar renewal equations (used in e.g.,[137,139]). These are a special case of our general framework (with $C = 2$ compartments). The conventional renewal equations additionally assume a constant generation time distribution over calendar time. Until now, the 1-D generation time distribution described by $w(\tau)$ has been included in frameworks directly as a function of the scalar delay τ (time from primary human infection to secondary human infection) where τ has not been decomposed into the times-since-state-entry variables and their corresponding joint density (as specified by g).

Monte Carlo simulation

Now, we demonstrate how to evaluate the 2-D integral in eq. (6.2.13) via Monte Carlo simulation. The simulation procedure simply requires samples (a_M, a_E) from $g(t, a_M, a_E)$, then $\tau^i = a_M^i + a_E^i$. This is equivalent to obtaining samples τ from $w(t, \tau)$, as defined in eq. (E.2.18). This procedure yields

$$\begin{aligned}
\int_0^t \int_0^{t-a_M} g(t, a_M, a_E) E(t - a_M - a_E, 0) da_E da_M &= \mathbb{E}_g[E(t - a_M - a_E, 0)] \\
&= \mathbb{E}_g[E(t-\tau, 0)] \\
&\approx \frac{1}{N} \sum_{i=1}^{N} E(t - \tau^i, 0)
\end{aligned}$$

where $\mathbb{E}_g$ denotes expectation with respect to the probability density for the (time-varying) generation time $g(t, a_M, a_E)$. We use the same Monte Carlo sampling procedure for our full four-compartment model (see Section 6.2.2) when deriving the density for our generation time over calendar time.

E.2.4 Analytical solutions for R_0 and r

To derive analytical solutions for $R(t)$ and r_t, we make the simplifying assumption that $b_{EM}(t, a_E) = e^{-\lambda_E a_E}$ and $b_{ME}(t, a_M) = e^{-\lambda_M a_M}$ which assumes that (for positive rate parameters) the birth processes decline with time since infection. As we have time-independent birth processes, we are actually deriving a basic reproduction number R_0 and intrinsic growth rate r where we take $t \to \infty$ in eq. (6.2.8) for the upper integral limits. So, eq. (6.2.8) becomes

$$\begin{aligned} 1 &= \phi(\gamma) \\ &= \int_0^\infty e^{-(\mu+\lambda_M+\gamma)a_M} \int_0^\infty e^{-(\lambda_E+\gamma)a_E} da_E da_M \\ &= \frac{1}{\lambda_E+\gamma} \int_0^\infty e^{-(\mu+\lambda_M+\gamma)a_M} da_M \\ &= \frac{1}{\lambda_E+\gamma} \cdot \frac{1}{\lambda_M+\gamma+\mu} \end{aligned} \tag{E.2.22}$$

So, r is the unique real root of eq. (E.2.22), and our basic reproduction number R_0 is the critical threshold for epidemic growth defined as

$$R_0 := \phi(0) = \frac{1}{\lambda_E} \cdot \frac{1}{\lambda_M+\mu} \quad . \tag{E.2.23}$$

For understanding the relationship between R_0 and r, we can use eq. (6.2.11) together with our previous assumptions. Then,

$$\begin{aligned} \frac{1}{R_0} &= \frac{\int_0^\infty e^{-(\mu+\lambda_M+r)a_M} \int_0^\infty e^{-(\lambda_E+r)a_E} da_E da_M}{\int_0^\infty e^{-(\mu+\lambda_M)a_M} \int_0^\infty e^{-\lambda_E a_E} da_E da_M} \\ &= \frac{\lambda_E(\lambda_M+\mu)}{(\lambda_E+r)(\lambda_M+\mu+r)} \end{aligned} \tag{E.2.24}$$

So, R_0 can also be written as a function of r (together with λ_E, λ_M, and μ) as follows:

$$R_0 = \frac{(\lambda_E+r)(\lambda_M+\mu+r)}{\lambda_E(\lambda_M+\mu)}. \tag{E.2.25}$$

E.3 Exposure-and-infection-age-structured human and infection-age-structured mosquito population

E.3.1 Modelling framework

We now describe an intermediate framework (with three compartments) between the frameworks described in Section 6.2.1 and Section 6.2.2. Building on the framework in Section 6.2.1, we incorporate separate exposed and infectious human compartments (E and I respectively), which are again structured by times since exposure and infection respectively. The infectious mosquito population is also structured by time since infection. The framework is described by the following conservation equations:

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a_E} = -b_{EI}\,(t, a_E)\,E\,(t, a_E) \tag{E.3.1}$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a_I} = 0 \tag{E.3.2}$$

$$\frac{\partial M}{\partial t} + \frac{\partial M}{\partial a_M} = -\mu M\,(t, a_M) \tag{E.3.3}$$

We close the system with the following initial and boundary conditions:

$$E(t,0) = \int_0^\infty b_{ME}\left(t,a_M\right) M\left(t,a_M\right) da_M \tag{E.3.4}$$

$$E\left(0,a_E\right) = f_E\left(a_E\right) \tag{E.3.5}$$

$$I(t,0) = \int_0^\infty b_{EI}\left(t,a_E\right) E\left(t,a_E\right) da_E \tag{E.3.6}$$

$$I\left(0,a_I\right) = f_I\left(a_I\right) \tag{E.3.7}$$

$$M(t,0) = \int_0^\infty b_{IM}\left(t,a_I\right) I\left(t,a_I\right) da_I \tag{E.3.8}$$

$$M\left(0,a_M\right) = f_M\left(a_M\right) \tag{E.3.9}$$

$$\tag{E.3.10}$$

where $\mu \geq 0$ is the infectious mosquito death rate (constant across time and ages), f_E, f_I and f_M denote the initial densities for the exposed human, infectious human, and infectious mosquito populations respectively, and each $b_{AB}(t,a_A) \geq 0$ is a time- and age-dependent birth function that represents the rate at which new individuals in sub-population B are generated at calendar time t by an individual of age a_A in sub-population A.

E.3.2 Deriving a renewal equation and instantaneous reproduction number

We first consider the characteristics for the mosquito population (for which there is no distinction between exposed or infectious states):

$$M(t,a_M) = \begin{cases} M(t-a_M,0)e^{-\mu a_M}, & \text{if } a_M \leq t, \\ f_M(a_M-t)e^{-\mu t}, & \text{if } a_M > t. \end{cases} \tag{E.3.11}$$

Then, we can rewrite (E.3.4) as:

$$\begin{aligned} E(t,0) = \int_0^t b_{ME}\left(t,a_M\right) M(t-a_M,0)e^{-\mu a_M} da_M + \\ \int_t^\infty b_{ME}\left(t,a_M\right) f_M(a_M-t)e^{-\mu t} da_M \end{aligned} \tag{E.3.12}$$

Again, since we are mostly concerned with the long-time behaviour of the system, i.e., when $t \gg a_M$ and $f_E(a_E-t)e^{-\mu t} \approx 0$. Then we can neglect the second part of eq. (E.3.12) and rewrite it as:

$$E(t,0) = \int_0^t b_{ME}\left(t,a_M\right) M(t-a_M,0)e^{-\mu a_M} da_M \tag{E.3.13}$$

Now using eq. (E.3.8), we have:

$$E(t,0) = \int_0^t b_{ME}\left(t,a_M\right) \int_0^\infty b_{IM}\left(t-a_M,a_I\right) I\left(t-a_M,a_I\right) da_I e^{-\mu a_M} da_M \tag{E.3.14}$$

Upon writing the characteristics of the infectious sub-population, for each of the two cases, we have the solution:

$$I(t,a_I) = \begin{cases} I(t-a_I,0), & \text{if } a_I \leq t, \\ f_I(a_I-t), & \text{if } a_I > t, \end{cases} \tag{E.3.15}$$

Considering the long-term dynamics of the solution, eq. (E.3.14) can be rewritten as

$$E(t,0) = \int_0^t b_{ME}(t,a_M) \int_0^{t-a_M} b_{IM}(t-a_M,a_I)\, I(t-a_M-a_I,0)\, da_I e^{-\mu a_M} da_M \tag{E.3.16}$$

Using (E.3.6), we have:

$$\begin{aligned} E(t,0) = &\int_0^t b_{ME}(t,a_M)\, e^{-\mu a_M} \int_0^{t-a_M} b_{IM}(t-a_M,a_I) \\ &\cdot \int_0^{\infty} b_{EI}(t-a_M-a_I,a_E)\, E(t-a_M-a_I,a_E)\, da_E da_I da_M \end{aligned} \tag{E.3.17}$$

Now, in a similar approach to that taken in Section 6.2.1, we derive a more conventional form of the renewal equation for the system. The characteristic of the exposed human sub-population yields

$$E(t,a_E) = \begin{cases} E(t-a_E,0)e^{-\int_0^{a_E} b_{EI}(t-a_E+s,s)ds}, & \text{if } a_E \le t \\ f_E(a_E-t)e^{-\int_{a_{E^{-t}}}^{a_E} b_{EI}(t,s)ds}, & \text{if } a_E > t \end{cases} \tag{E.3.18}$$

Focusing on long-term dynamics (i.e., $t \gg a_E$ and hence $f_E(a_E-t)e^{-\int_{a_{E^{-t}}}^{a_E} b_{EI}(t,s)ds} = 0$), we have the following renewal equation (where the incidence of new infections is a function of the new infections at earlier times):

$$\begin{aligned} E(t,0) = &\int_0^t b_{ME}(t,a_M)\, e^{-\mu a_M} \int_0^{t-a_M} b_{IM}(t-a_M,a_I) \\ &\cdot \int_0^{t-a_M-a_I} b_{EI}(t-a_M-a_I,a_E)\, E(t-a_M-a_I-a_E,0)\, e^{-\int_0^{a_E} b_{EI}(t-a_M-a_I-a_E+s,s)ds} da_E da_I da_M \end{aligned} \tag{E.3.19}$$

As we are primarily considered with a long-term solution of our system, similar to Section (6.2.1), we derive a solution for the exposed human compartment of the system of the form:

$$E(t,a_E) = p(a_E)e^{\gamma t}. \tag{E.3.20}$$

This solution states that the age distribution of exposed individuals is altered by a factor which grows or decays with time depending on whether $\gamma > 0$ or $\gamma < 0$ respectively. So, if we are to assume such a separable solution (to derive analytical results), then we need that $b_{EI}(t,a_E) = b_{EI}(a_E)\,\forall t$. So, we assume that the (per capita) rate at which an exposed individual becomes infectious is independent of calendar time and depends only on time since exposure.

We first substitute the expression from eq. (E.3.20) into eq. (E.3.1) to obtain:

$$\gamma p(a_E)e^{\gamma t} + \frac{dp(a_E)}{da_E}e^{\gamma t} = -b_{EI}(a_E)\, p(a_E)e^{\gamma t}, \tag{E.3.21}$$

which implies

$$p(a_E) = p(0)e^{-\gamma\left(a_E+\int_0^{a_E} b_{EI}(s)ds\right)}. \tag{E.3.22}$$

Substituting the long-term solution and eq. (E.3.22) into eq. (E.3.19) yields:

$$\begin{aligned} p(0)e^{\gamma t} = &\int_0^t b_{ME}(t,a_M)\, e^{-\mu a_M} \int_0^{t-a_M} b_{IM}(t-a_M,a_I) \\ &\cdot \int_0^{t-a_M-a_I} b_{EI}(a_E)\, e^{-\int_0^{a_E} b_{EI}(s)ds} p(0)e^{\gamma(t-a_M-a_I-a_E)} da_E da_I da_M \end{aligned} \tag{E.3.23}$$

After cancelling $p(0)e^{\gamma t}$ from both sides of the equation, we have:

$$\begin{aligned} 1 = &\int_0^t b_{ME}(t,a_M)\, e^{-\mu a_M} \int_0^{t-a_M} b_{IM}(t-a_M,a_I) \\ &\cdot \int_0^{t-a_M-a_I} b_{EI}(a_E)\, e^{-\int_0^{a_E} b_{EI}(s)ds} e^{-\gamma(a_M+a_I+a_E)} da_E da_I da_M =: \phi(\gamma) \end{aligned} \tag{E.3.24}$$

Now, as $\phi(\gamma)$ is again a monotonically decreasing function of γ, identical to the system in Section 6.2.1, we consider $\phi(0)$ to derive a reproduction number formula:

$$\begin{aligned} R(t) = \int_0^t b_{ME}\left(t, a_M\right) e^{-\mu a_M} \int_0^{t-a_M} b_{IM}\left(t-a_M, a_I\right) \\ \cdot \int_0^{t-a_M-a_I} b_{EI}\left(a_E\right) e^{-\int_0^{a_E} b_{EI}(s)ds} da_E da_I da_M \end{aligned} \tag{E.3.25}$$

Let $R_{EI}(t) := \int_0^t b_{EI}\left(a_E\right) e^{-\int_0^{a_E} b_{EI}(s)ds} da_E$, which again can be used to derive a normalised version of the equation: $1 = \int_0^t \beta_{EI}(a_E) da_E$ where $\beta_{EI}(a_E) := b_{EI}\left(a_E\right) e^{-\int_0^{a_E} b_{EI}(s)ds} / R_{EI}(t)$.

Replicating a similar procedure for the transitions between the mosquito and exposed human sub-populations, we define $R_{ME}(t) := \int_0^t b_{ME}(t, a_M) e^{-\mu a_M} da_M$, which leads to the equation: $1 = \int_0^t \beta_{ME}(t, a_M) da_M$ where $\beta_{ME}(t, a_M) := b_{ME}(t, a_M) e^{-\mu a_M} / R_{ME}(t)$.

We can also write $R_{IM}(t) := \int_0^t b_{IM}(t, a_I) da_I$, which leads to the equation: $1 = \int_0^t \beta_{IM}(t, a_I) da_I$ where $\beta_{IM}(t, a_I) := b_{IM}(t, a_I) / R_{IM}(t)$.

This suggests writing eq. (E.3.25) as:

$$\begin{aligned} R(t) &= R_{ME}(t) \mathbb{E}_{ME}\left[R_{IM}(t-a_M) \int_0^{t-a_M} \beta_{IM}\left(t-a_M, a_I\right) R_{EI}\left(t-a_M-a_I\right) da_I\right] \\ &= R_{ME}(t) \mathbb{E}_{ME}\left[R_{IM}(t-a_M) \mathbb{E}_{IM}\left[R_{EI}\left(t-a_M-a_I\right)\right]\right] \end{aligned} \tag{E.3.26}$$

where the expectations are defined as $\mathbb{E}_{ME}[X(t-a_M)] := \int_0^t X(t-a_M) \beta_{ME}(t, a_M) da_M$ and $\mathbb{E}_{IM}[X(t-a_M-a_I)] := \int_0^{t-a_M} X(t-a_M-a_I) \beta_{IM}(t-a_M, a_I) da_I$.

Alternatively, we have

$$\begin{aligned} R(t) &= \int_0^t \int_0^{t-a_M} \int_0^{t-a_M-a_I} b_{ME}\left(t, a_M\right) e^{-\mu a_M} b_{IM}\left(t-a_M, a_I\right) b_{EI}\left(a_E\right) \\ &\quad \cdot e^{-\int_0^{a_E} b_{EI}(s)ds} da_E da_I da_M \\ &= \int_0^t \int_0^{t-a_M} \int_0^{t-a_M-a_I} K(t, a_E, a_I, a_M) da_E da_I da_M \end{aligned} \tag{E.3.27}$$

which suggests definition of a probability density for the generation time at time t,

$$g(t, a_E, a_I, a_M) := K(t, a_E, a_I, a_M) / R(t). \tag{E.3.28}$$

We can write eq. (E.3.19) as:

$$E(t, 0) = R(t) \int_0^t \int_0^{t-a_M} \int_0^{t-a_M-a_I} g(t, a_E, a_I, a_M) E\left(t-a_M-a_I-a_E, 0\right) da_E da_I da_M \tag{E.3.29}$$

E.3.3 Relation to the simple ubiquitous renewal equation frameworks

Similar to Section 6.2.1, we can relate our derived renewal equation is to the renewal equations frequently used in directly transmitted infectious diseases.

Let $\tau = a_E + a_I + a_M$ denote the total delay from primary human infection to secondary transmission, derived by adding the times-between-states (i.e., the stages of the transmission cycle).

Focusing on the integral

$$\int_0^t \int_0^{t-a_M} \int_0^{t-a_M-a_I} g(t, a_E, a_I, a_M) E\left(t - a_M - a_I - a_E, 0\right) da_E da_I da_M$$

from eq. (E.3.29), we will now use a change of variables to convert $(a_E, a_I, a_M) \rightarrow (a_E, a_I, \tau)$. Using bijective functions $l(a_E, a_I, \tau) = a_E$, $m(a_E, a_I, \tau) = a_I$ and $n(a_E, a_I, \tau) = a_M = \tau - a_E - a_I$, we have:

$$\begin{aligned}
&\int_0^t \int_0^{t-a_M} \int_0^{t-a_M-a_I} g(t, a_E, a_I, a_M) E\left(t - a_M - a_I - a_E, 0\right) da_E da_I da_M \\
&= \int_0^t \int_0^{\tau} \int_0^{\tau-a_I} g(t, l(a_E, a_I, \tau, m(a_E, a_I, \tau), n(a_E, a_I, \tau)) \\
&\cdot E(t - n(a_E, a_I, \tau) - m(a_E, a_I, \tau) - l(a_E, a_I, \tau), 0) \left|\frac{\partial(a_E, a_I, a_M)}{\partial(a_E, a_I, \tau)}\right| da_E da_I d\tau \\
&= \int_0^t \int_0^{\tau} \int_0^{\tau-a_I} g(t, a_E, a_I, \tau - a_E - a_I) E(t - a_E - a_I - (\tau - a_E - a_I), 0) da_E da_I d\tau \\
&= \int_0^t E(t - \tau, 0) \int_0^{\tau} \int_0^{\tau-a_I} g(t, a_E, a_I, \tau - a_E - a_I) da_E da_I d\tau \\
&= \int_0^t E(t - \tau, 0) \int_0^{\tau} [\tilde{g}(t, a_E, a_I, \tau - a_E - a_I)]_0^{\tau-a_I} da_I d\tau \\
&= \int_0^t E(t - \tau, 0) w(t, \tau)\, d\tau
\end{aligned} \tag{E.3.30}$$

where $\left|\frac{\partial(a_E, a_I, a_M)}{\partial(a_E, a_I, \tau)}\right| = 1$ denotes the Jacobian (i.e., the determinant of the Jacobian matrix), $\tilde{g}(t, a_E, a_I, \tau - a_E - a_I)$ is the antiderivative of $g(t, a_E, a_I, \tau - a_E - a_I)$ with respect to a_E, and $w(t, \tau)$ is a definite integral (of a definite integral of $\tilde{g}$ with respect to a_I) defined as $w(t, \tau) := \int_0^{\tau} [\tilde{g}(t, a_E, a_I, \tau - a_E - a_I)]_0^{\tau-a_I} da_I$. Here, $[\tilde{g}(t, a_E, a_I, \tau - a_E - a_I)]_0^{\tau-a_I}$ is a function of (t, a_I, τ) as the dependence on a_E has been integrated (i.e., marginalised) out from the expression.

Similar to eq. (E.2.19), we can relate eq. (E.3.29) and hence our framework to the ubiquitous renewal equation framework of directly transmitted infectious diseases, albeit with a time-varying generation time $w(t, \tau)$:

$$E(t, 0) = R(t) \int_0^t E(t - \tau, 0) w(t, \tau) d\tau. \tag{E.3.31}$$

In short, the relation (and compression to a 1-D marginal density $w(t, \tau)$ at time t), has been established by i) the Jacobian-based change of variable (to frame our integral equation in terms of the sum of the times between states) and ii) $C - 1 = 2$ (inner) integration steps (where $C = 3$ is the number of model compartments) to marginalise out the first two times-between state variables.

Once again, assuming a constant density for the generation time over calendar time $g(t, a_E, a_I, a_M) = g(a_E, a_I, a_M) \forall t$, we arrive at the ubiquitous renewal equation framework of directly transmitted infectious diseases:

$$E(t, 0) = R(t) \int_0^t E(t - \tau, 0) w(\tau) d\tau. \tag{E.3.32}$$

Here, assuming a constant generation time over calendar time is equivalent to assuming that the relative contribution of each times-between-state triplet (a_E, a_I, a_M) to secondary human infections is fixed throughout the course of an outbreak.

In further detail, this means that the numerator (the instantaneous kernel at time t) and denominator (the reproduction number at time t) of the following expression always appear in fixed proportions:

$$\begin{aligned} g(a_E, a_I, a_M) &= \frac{K(t, a_E, a_I, a_M)}{R(t)} \\ &= \frac{b_{ME}\,(t, a_M)\, e^{-\mu a_M} b_{IM}\,(t - a_M, a_I)\, b_{EI}\,(a_E)\, e^{-\int_0^{a_E} b_{EI}(s)ds}}{\int_0^t \int_0^{t-a_M} \int_0^{t-a_M-a_I} K(t, a_E, a_I, a_M) da_E da_I da_M}. \end{aligned} \tag{E.3.33}$$

Again, as we have described in detail in Section 6.2.1, the assumption of calendar-time-homogeneous relative contributions of times-in-state triplet (a_E, a_I, a_M) may be unrealistic for transmission of vector-borne diseases. Nevertheless, via eq. (E.3.32), we have formalised a relationship between our derived framework and the more conventional renewal equation frameworks.

E.4 Exposure-and-infection-age-structured populations of human and mosquitoes

E.4.1 Modelling framework

We now provide further details of the framework described in Section 6.2.2. Recall that the system was described with the following linear PDEs for the exposed (E) and infectious (I) human populations, and the exposed (W) and infectious (V) mosquito populations:

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a_E} = -b_{EI}\left(t, a_E\right) E\left(t, a_E\right) \tag{E.4.1}$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a_I} = 0 \tag{E.4.2}$$

$$\frac{\partial W}{\partial t} + \frac{\partial W}{\partial a_W} = -b_{WV}\left(t, a_W\right) W\left(t, a_W\right) - \mu W\left(t, a_W\right) \tag{E.4.3}$$

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial a_V} = -\mu V\left(t, a_V\right) \tag{E.4.4}$$

where again each $b_{AB}(t, a_A) \geq 0$ is a time- and age-dependent birth function that represents the rate at which new individuals in sub-population B are generated at calendar time t by an individual of age a_A in sub-population A. $\mu \geq 0$ is the mosquito death rate for exposed and infectious mosquitoes, which we assume is age-independent[608].

We close the system with the following initial and boundary conditions:

$$E(t, 0) = \int_0^\infty b_{VE}\left(t, a_V\right) V\left(t, a_V\right) da_V \tag{E.4.5}$$

$$E\left(0, a_E\right) = f_E\left(a_E\right) \tag{E.4.6}$$

$$I(t, 0) = \int_0^\infty b_{EI}\left(t, a_E\right) E\left(t, a_E\right) da_E \tag{E.4.7}$$

$$I\left(0, a_I\right) = f_I\left(a_I\right) \tag{E.4.8}$$

$$W(t, 0) = \int_0^\infty b_{IW}\left(t, a_I\right) I\left(t, a_I\right) da_I \tag{E.4.9}$$

$$W\left(0, a_W\right) = f_W\left(a_W\right) \tag{E.4.10}$$

$$V(t, 0) = \int_0^\infty b_{WV}\left(t, a_W\right) W\left(t, a_W\right) da_W \tag{E.4.11}$$

$$V\left(0, a_V\right) = f_V\left(a_V\right) \tag{E.4.12}$$

where $f_E(a_E)$, $f_I(a_I)$, $f_W(a_W)$, and $f_V(a_V)$ are the initial densities for the exposed human, infectious human, exposed mosquito, and infectious mosquito sub-populations.

For this system, from the characteristics for the following sub-populations, we have:

$$E(t,a_E) = \begin{cases} E(t-a_E,0)e^{-\int_0^{a_E} b_{EI}(t-a_E+s,s)ds}, & \text{if } a_E \le t \\ f_E(a_E-t)e^{-\int_{a_E-t}^{a_E} b_{EI}(t,s)ds}, & \text{if } a_E > t \end{cases} \tag{E.4.13}$$

$$I(t,a_I) = \begin{cases} I(t-a_I,0), & \text{if } a_I \le t \\ f_I(a_I-t), & \text{if } a_I > t \end{cases} \tag{E.4.14}$$

$$W(t,a_W) = \begin{cases} W(t-a_W,0)\, e^{-\int_0^{a_W} b_{WV}(t-a_W+r,r)dr-\mu a_W}, & \text{if } a_W \le t \\ f_W(a_W-t)\, e^{-\int_{a_W-t}^{a_W} b_{WV}(t,r)dr-\mu t}, & \text{if } a_W > t \end{cases} \tag{E.4.15}$$

$$V(t,a_V) = \begin{cases} V(t-a_V,0)\, e^{-\mu a_V}, & \text{if } a_V \le t \\ f_V(a_V-t)\, e^{-\mu t}, & \text{if } a_V > t \end{cases} \tag{E.4.16}$$

E.4.2 Deriving a renewal equation and time-varying reproduction number

We now provide further details of the derivation for our instantaneous reproduction number, eq. (6.2.34). Once again, we model the length of time between primary human infection and secondary human transmission, i.e., the interval between a newly exposed individual and the secondary exposed human originating from this infection. Rewriting eq. (E.4.5), we have:

$$\begin{aligned} E(t,0) = &\int_0^t b_{VE}(t,a_V)\, V(t-a_V,0)\, e^{-\mu a_V} da_V \\ &+ \int_t^\infty b_{VE}(t,a_V)\, f_V(a_V-t)\, e^{-\mu t} da_V \end{aligned} \tag{E.4.17}$$

Once again, we focus primarily on the long-term behaviour of the system, i.e., when $t \gg a_V$ and $f_V(a_V-t)\, e^{-\mu t} \approx 0$. We can ignore the second part of the equation and also use eq. (E.4.7) to write:

$$E(t,0) = \int_0^t b_{VE}(t,a_V)\, e^{-\mu a_V} \int_0^\infty b_{WV}(t-a_V,a_W)\, W(t-a_V,a_W)\, da_W da_V \tag{E.4.18}$$

Using the characteristics for the exposed mosquito population (eq. (E.4.15)), and again focusing on $t \gg a_W$ (such that $f_W(a_W-t)\, e^{-\int_{a_W-t}^{a_W} b_{WV}(t,r)dr-\mu t} \approx 0$) we have:

$$\begin{aligned} E(t,0) = &\int_0^t b_{VE}(t,a_V)\, e^{-\mu a_V} \int_0^{t-a_V} b_{WV}(t-a_V,a_W)\, W(t-a_V-a_W,0) \\ &\cdot e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} da_W da_V \end{aligned} \tag{E.4.19}$$

Substituting in eq. (E.4.9), we have:

$$\begin{aligned} E(t,0) = &\int_0^t b_{VE}(t,a_V)\, e^{-\mu a_V} \int_0^{t-a_V} b_{WV}(t-a_V,a_W) \int_0^\infty b_{IW}(t-a_V-a_W,a_I) \\ &\cdot I(t-a_V-a_W,a_I)\, e^{-\int_0^{a_W} b_{WV}(t-a_V,r)dr-\mu a_W} da_I da_W da_V \end{aligned} \tag{E.4.20}$$

Now, using the characteristics for the infectious human population and focusing on long-term dynamics, we have:

$$\begin{aligned} E(t,0) = &\int_0^t b_{VE}(t,a_V)\, e^{-\mu a_V} \int_0^{t-a_V} b_{WV}(t-a_V,a_W)\, e^{-\int_0^{a_W} b_{WV}(t-a_V,r)dr-\mu a_W} \\ &\cdot \int_0^{t-a_V-a_W} b_{IW}(t-a_V-a_W,a_I)\, I(t-a_V-a_W-a_I,0)\, da_I da_W da_V \end{aligned} \tag{E.4.21}$$

Using the boundary condition for the infectious human population, we have:

$$\begin{aligned} E(t,0) = &\int_0^t b_{VE}(t,a_V)\, e^{-\mu a_V} \int_0^{t-a_V} b_{WV}(t-a_V,a_W) \int_0^{t-a_V-a_W} b_{IW}(t-a_V-a_W,a_I) \\ &\cdot e^{-\int_0^{a_W} b_{WV}(t-a_V,r)dr-\mu a_W} \int_0^{\infty} b_{EI}(t-a_V-a_W-a_I,a_E)\, E(t-a_V-a_W-a_I,a_E) \\ &\cdot da_E da_I da_W da_V \end{aligned} \tag{E.4.22}$$

We can rearrange eq. (E.4.22) to emphasise the role of the birth processes and relatedly, the ages of the source individuals at the times at which secondary births occur:

$$\begin{aligned} E(t,0) = &\int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{\infty} b_{VE}(t,a_V)\, b_{WV}(t-a_V,a_W)\, b_{IW}(t-a_V-a_W,a_I) \\ &\cdot b_{EI}(t-a_V-a_W-a_I,a_E)\, e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V,r)dr-\mu a_W} E(t-a_V-a_W-a_I,a_E) \\ &\cdot da_E da_I da_W da_V \end{aligned} \tag{E.4.23}$$

Using eq. (E.4.13), focusing on long-term dynamics ($t \gg a_E$), we have the following renewal equation:

$$\begin{aligned} E(t,0) = &\int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} b_{VE}(t,a_V)\, b_{WV}(t-a_V,a_W)\, b_{IW}(t-a_V-a_W,a_I) \\ &\cdot b_{EI}(t-a_V-a_W-a_I,a_E)\, e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} \\ &\cdot E(t-a_V-a_W-a_I-a_E,0) e^{-\int_0^{a_E} b_{EI}(t-a_V-a_W-a_I-a_E+s,s)ds} \\ &\cdot da_E da_I da_W da_V \end{aligned} \tag{E.4.24}$$

We are primarily considered with a long-term solution of our system, and now, similar to Section (E.3), we derive a solution for the infectious human compartment of the system of the form:

$$E(t,a_E) = p(a_E)e^{\gamma t}. \tag{E.4.25}$$

This separable solution amounts to assuming that the age-since-exposure distribution of the exposed human compartment, is altered by a time-dependent factor which either grows ($\gamma > 0$) or decays ($\gamma < 0$). So, if we are to assume such a separable solution (to derive analytical results), then we need $b_{EI}(t,a_E) = b_{EI}(a_E)\,\forall t$. So, we assume that the (per capita) rate at which an exposed individual becomes infectious is independent of calendar time and depends only on time since exposure.

We first substitute the expression from eq.(E.4.25) into eq. (E.4.1)to obtain:

$$\gamma p(a_E)e^{\gamma t} + \frac{dp(a_E)}{da_E}e^{\gamma t} = -b_{EI}(a_E)\, p(a_E)e^{\gamma t}, \tag{E.4.26}$$

which implies

$$p(a_E) = p(0)e^{-\gamma\left(a_E+\int_0^{a_E} b_{EI}(s)ds\right)}. \tag{E.4.27}$$

Substituting the separable solution (eq. (E.4.25)) and eq. (E.4.27) into eq. (E.4.24) yields:

$$\begin{aligned} p(0)e^{\gamma t} = &\int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} b_{VE}(t,a_V)\, b_{WV}(t-a_V,a_W)\, b_{IW}(t-a_V-a_W,a_I) \\ &\cdot b_{EI}(a_E)\, e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} e^{-\int_0^{a_E} b_{EI}(s)ds} \\ &\cdot p(0)e^{\gamma(t-a_V-a_W-a_I-a_E)} da_E da_I da_W da_V \end{aligned} \tag{E.4.28}$$

where we have used the assumption that $b_{EI}(t,a_E) = b_{EI}(a_E)\,\forall t$.

Upon cancelling $p(0)e^{\gamma t}$ from both sides of the equation, we have:

$$\begin{aligned} 1 = &\int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} b_{VE}(t,a_V)\, b_{WV}(t-a_V,a_W)\, b_{IW}(t-a_V-a_W,a_I) \\ &\cdot b_{EI}(a_E)\, e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} e^{-\int_0^{a_E} b_{EI}(s)ds} \\ &\cdot e^{-\gamma(a_V+a_W+a_I+a_E)} da_E da_I da_W da_V =: \phi(\gamma) \end{aligned} \tag{E.4.29}$$

Now, as $\phi(\gamma)$ is a decreasing function of γ, similar to the system in Section 6.2.1, we have that the exponential growth/decay of human infections is controlled by the growth/decay parameter γ alongside times spent in each of the four individual compartments.

We consider $\phi(0)$ to derive a reproduction number formula:

$$\begin{aligned} R(t) = &\int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} b_{VE}(t,a_V)\, b_{WV}(t-a_V,a_W)\, b_{IW}(t-a_V-a_W,a_I) \\ &\cdot b_{EI}(a_E)\, e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} \\ &\cdot e^{-\int_0^{a_E} b_{EI}(s)ds} da_E da_I da_W da_V \end{aligned} \tag{E.4.30}$$

Alternatively, we can write:

$$\begin{aligned} R(t) = &\int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} b_{VE}(t,a_V)\, b_{WV}(t-a_V,a_W)\, b_{IW}(t-a_V-a_W,a_I) \\ &\cdot b_{EI}(a_E)\, e^{-\mu a_V} e^{-\int_0^{a_W} b_{WV}(t-a_V-a_W+r,r)dr-\mu a_W} \\ &\cdot e^{-\int_0^{a_E} b_{EI}(s)ds} da_E da_I da_W da_V \\ = &\int_0^t \int_0^{t-a_V} \int_0^{t-a_V-a_W} \int_0^{t-a_V-a_W-a_I} K(t,a_E,a_I,a_W,a_V) da_E da_I da_W da_V \end{aligned} \tag{E.4.31}$$

which now suggests definition of a four-dimensional probability density $g(t,a_E,a_I,a_W,a_V) := K(t,a_E,a_I,a_W,a_V)/R(t)$. As our instantaneous reproduction number defines the expected number of secondary human infections produced by an individual initially infected at time t (if conditions remained constant), K denotes the number of secondary human infections are generated at calendar time t for various individual times spent (represented as ages) in each state/compartment, whilst g denotes the normalised kernel (otherwise known as the instantaneous generation time[145]) which represents the relative contribution of past human infections, previously exposed humans, previously exposed mosquitoes, and currently infectious mosquitoes (across their various times within compartments i.e., ages) to the current number of humans currently infectious. In each case, for either g, K, or R, we are describing a counterfactual scenario of contributions to the epidemic where the conditions would remain constant.

E.4.3 Between-stage reproduction numbers

As before, focusing on generation events between stages, we define

$$\begin{aligned} R_{VE}(t) &:= \int_0^t b_{VE}\,(t, a_V)\, e^{-\mu a_V} da_V, \\ \beta_{VE}(a_V) &:= \frac{b_{VE}\,(t, a_V)\, e^{-\mu a_V}}{R_{VE}(t)}, \end{aligned} \tag{E.4.32}$$

where $1 = \int_0^t \beta_{VE}(t, a_V) da_V$. Replicating similar procedures for other sub-populations, we also define

$$\begin{aligned} R_{IW}(t) &:= \int_0^t b_{IW}(t, a_I) da_I, \\ \beta_{IW}(t, a_I) &:= \frac{b_{IW}(t, a_I)}{R_{IW}(t)}, \end{aligned} \tag{E.4.33}$$

where $1 = \int_0^t \beta_{IW}(t, a_I) da_I$. Also, let

$$\begin{aligned} R_{WV}(t) &:= \int_0^t b_{WV}(t, a_W) e^{-\int_0^{a_W} b_{WV}(t-a_W+r, r) dr} da_W, \\ \beta_{WV}(t, a_W) &:= \frac{b_{WV}(t, a_W) e^{-\int_0^{a_W} b_{WV}(t-a_W+r, r) dr - \mu a_W}}{R_{WV}(t)}, \end{aligned} \tag{E.4.34}$$

where $1 = \int_0^t \beta_{WV}(t, a_W) da_W$. Finally, let

$$\begin{aligned} R_{EI}(t) &:= \int_0^t b_{EI}\,(a_E)\, e^{-\int_0^{a_E} b_{EI}(a_E) ds} da_E \\ \beta_{EI}(t, a_E) &:= \beta_{EI}(a_E) = \frac{b_{EI}\,(a_E)\, e^{-\int_0^{a_E} b_{EI}(a_E) ds}}{R_{EI}(t)} \end{aligned} \tag{E.4.35}$$

where $1 = \int_0^t \beta_{EI}(t, a_E) da_E$. So, we have

$$\begin{aligned} \beta_{EI}(t, a_E) = \beta_{EI}(a_E) &= \frac{K_{EI}(a_E)}{R_{EI}(t)}, \\ \beta_{IW}(t, a_I) = \frac{b_{IW}(t, a_I)}{R_{IW}(t)} &= \frac{K_{IW}(t, a_I)}{R_{IW}(t)}, \\ \beta_{WV}(t, a_W) = \frac{b_{WV}(t, a_W) e^{-\int_0^{a_W} b_{WV}(t-a_W+r, r) dr - \mu a_W}}{R_{WV}(t)} &= \frac{K_{WV}(t, a_W)}{R_{WV}(t)}, \\ \beta_{VE}(t, a_V) = \frac{b_{VE}\,(t, a_V)\, e^{-\mu a_V}}{R_{VE}(t)} &= \frac{K_{VE}(t, a_V)}{R_{VE}(t)}. \end{aligned} \tag{E.4.36}$$

Each $\beta_{CD}(t, a_C)$ defines a probability density (integrating to 1 at each calendar time t) for the generation time between stages (C and D). It is obtained by normalising the corresponding instantaneous kernels $K_{CD}(t, a_C)$. $\beta_{CD}(t, a_C)$ determines the distribution (at calendar time t) for the times taken from initial entry (into compartment C) to the generation of a new individual in the other compartment (D). The time-varying nature means that the relative contributions of different times-in-state (a_C) vary at different calendar times. This may happen in our setting of vector-borne diseases (e.g., under time variations caused by climatic conditions or changing totals of susceptible and infectious hosts and vectors). For example, we capture time variance in the extrinsic latent period over calendar time via $\beta_{WV}(t, a_W)$, measuring from initial mosquito exposure (W) to infectious mosquito state (V) the relative contribution of each age a_W (time-since-mosquito-exposure) to the generation of a newly infectious mosquito.

So, we can rewrite eq. (6.2.34) as

$$R(t) = R_{VE}(t) \int_0^t \beta_{VE}(t, a_V) R_{WV}(t - a_V) \int_0^{t-a_V} \beta_{WV}(t - a_V, a_W) R_{IW}(t - a_V - a_W) \cdot \int_0^{t-a_V-a_W} \beta_{IW}(t - a_V - a_W, a_I) R_{EI}(t - a_V - a_W - a_I) da_I da_W da_V$$

and using expectations defined as

$$\mathbb{E}_{VE}[X(t - a_V)] := \int_0^t X(t - a_V)\beta_{VE}(t, a_V) da_V,$$
$$\mathbb{E}_{WV}[X(t - a_V - a_W)] := \int_0^{t-a_V} X(t - a_V - a_W)\beta_{WV}(t - a_V, a_W) da_W,$$
$$\mathbb{E}_{IW}[X(t - a_V - a_W - a_I)] := \int_0^{t-a_V-a_W} X(t - a_V - a_W - a_I)\beta_{IW}(t - a_V - a_W, a_I) da_I,$$

we also have:

$$R(t) = R_{VE}(t)\mathbb{E}_{VE}\left[R_{WV}(t - a_V)\mathbb{E}_{WV}\left[R_{IW}(t - a_V - a_W)\mathbb{E}_{IW}\left[R_{EI}\left(t - a_V - a_W - a_I\right)\right]\right]\right]. \tag{E.4.37}$$

Again, the instantaneous reproduction number is an expectation of between-stage reproduction numbers across stages of the transmission cycle (the cycle required between two successive human infections). The between-stage reproduction numbers are evaluated with upper limits determined by calendar times at which the subsequent stage of the transmission cycle begins. The expectations are taken with respect to the generation time between stages of the next stage of the transmission cycle. The generation times are evaluated using ages of the sub-populations at the calendar times at which onward generation events in the transmission cycle occur (see Figure 6.1C: a_V at time t, a_W at time $t - a_V$, and a_I at time $t - a_V - a_W$).

E.4.4 Application to simulated and real-world data

We provide here further details of the Monte Carlo sampling, distributional assumptions, and results from the application in Section 6.2.2. The distributional assumptions were all based on existing work by Siraj et al. [156].

In Siraj et al. [156], the time from exposed (E) to infectious (I) was approximated by the intrinsic incubation period (the time from human infection to symptom onset). Based on results from a previous modelling review [80], the intrinsic incubation period was modelled using a log-normal distribution with mean of 5.97 days and standard deviation of 1.64 days. Denote this probability density as $\beta_{EI}(t, a_E)$.

For the time from infectious human (I) to infected mosquito (W), Siraj et al. [156] used a normal probability density function multiplied by a calibration/scaling factor to fit to previously published experimental data [765]. Denote this probability density as $\beta_{IW}(t, a_I)$.

Similar to the human population, the time from infected (W) to infectious (V) mosquito was approximated in Siraj et al. [156] using the extrinsic incubation period (EIP), based on results from the same modelling review [80]. This EIP was modelled with a log-normal distribution with a shape parameter of 4.9 and scale parameter of $e^{2.9-0.08 \cdot T}$ where T is the observed temperature. Denote this probability density as $\beta_{WV}(t, a_W)$.

The period from the infectious mosquito (V) to infected human (E) was modelled in Siraj et al. [156] using an existing temperature- and age-dependent model [81] with an additional rate of extrinsic mortality (of 0.089 per day). Denote this probability density as $\beta_{VE}(t, a_V)$.

So, for the existing approach, at each iteration i of our Monte Carlo sampling scheme, we can generate sample values $a_E^{(i)}, a_I^{(i)}, a_W^{(i)}$, and $a_V^{(i)}$ using the distributions described by $\beta_{EI}(t, a_E), \beta_{IW}(t, a_I), \beta_{WV}(t, a_W)$, and $\beta_{VE}(t, a_V)$. For the stage-specific approach, at each iteration i of our Monte Carlo sampling scheme, we generate sample values $a_V^{(i)}$

using $\beta_{VE}(t, a_V)$. Then, we sample $a_W^{(i)}$ using $\beta_{WV}(t - a_V^{(i)}, a_W)$, $a_I^{(i)}$ using $\beta_{IW}(t - a_V^{(i)} - a_W^{(i)}, a_W)$, and $a_E^{(i)}$ using $\beta_{EI}(t - a_V^{(i)} - a_W^{(i)} - a_I^{(i)}, a_E)$. For either approach (stage-specific or existing), at iteration i, we let $\tau^{(i)} = a_E^{(i)} + a_I^{(i)} + a_W^{(i)} + a_V^{(i)}$. Repeating this procedure for $i \in (1, \ldots, 2000)$ produces 2000 Monte Carlo samples to approximate the two-dimensional generation time distribution $w(t, \tau)$.

In addition to Figure 6.3, we present results for Ang Mo Kio, São João do Tauape, and the simulated environment in Figures SI E.1-SI E.6.

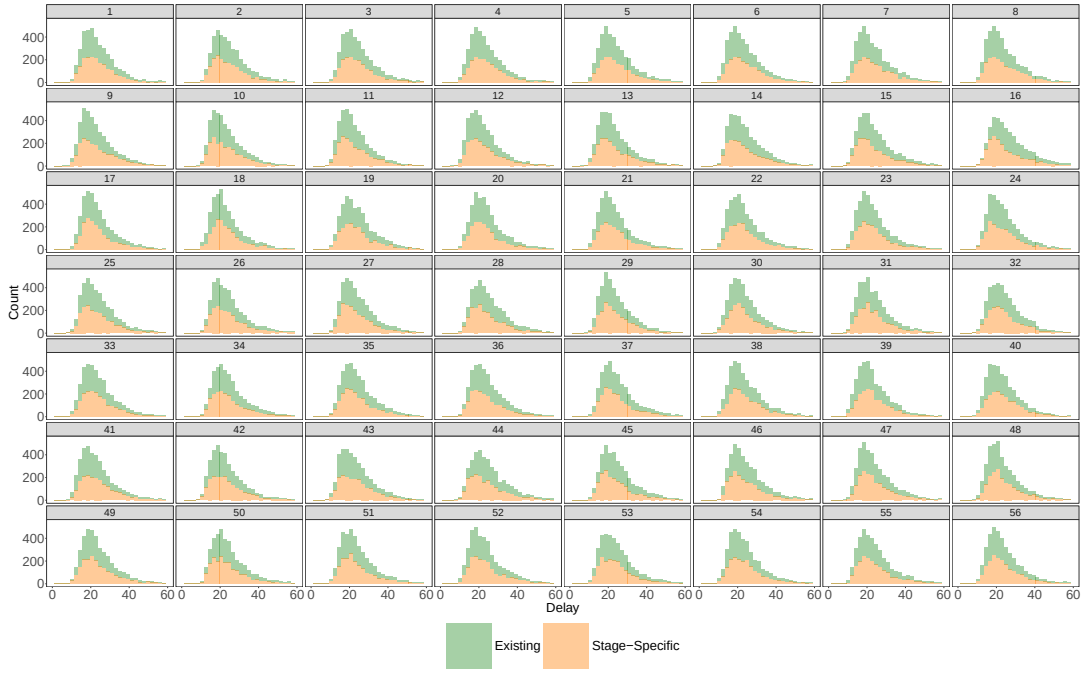


Figure SI E.1: **Time-varying generation times in Ang Mo Kio, Singapore**. For individual calendar times, represented by numbers in facet labels, we present histograms of Monte Carlo samples from the generation time distribution under different temperature-dependent assumptions. For the existing (light blue) approach, we allow the transmission cycle to change every day based on the temperature observed that day. Our stage-specific approach means that the generation time distribution can vary based on temperatures observed for calendar times when the estimated (by Monte Carlo samples) generation events occurred.

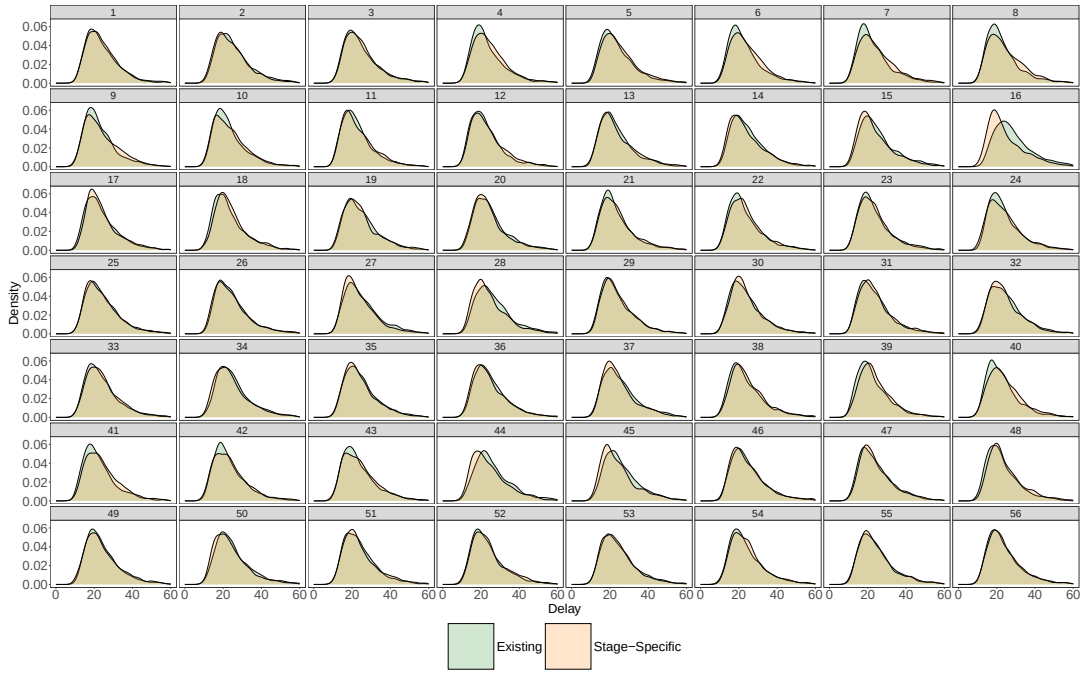


Figure SI E.2: **Time-varying generation times in Ang Mo Kio, Singapore**. For individual calendar times, represented by numbers in facet labels, we present kernel density estimates of Monte Carlo samples from the generation time distribution under different temperature-dependent assumptions. For the existing (light blue) approach, we allow the transmission cycle to change every day based on the temperature observed that day. Our stage-specific approach means that the generation time distribution can vary based on temperatures observed for calendar times when the estimated (by Monte Carlo samples) generation events occurred.

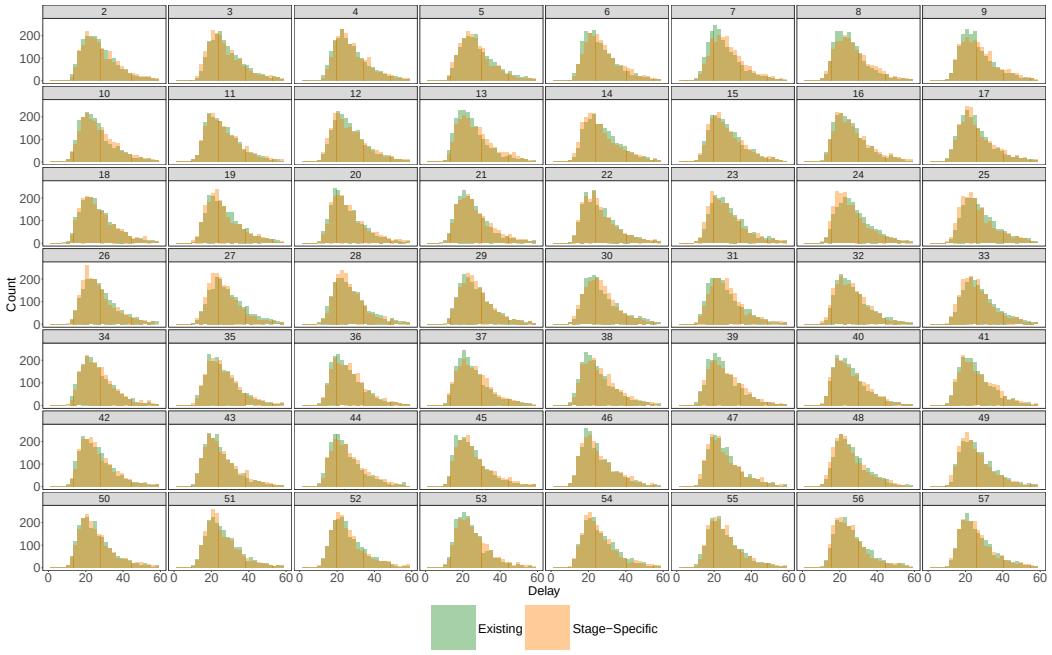


Figure SI E.3: **Time-varying generation times in São João do Tauape, Fortaleza, Brazil**. For individual calendar times, represented by numbers in facet labels, we present histograms of Monte Carlo samples from the generation time distribution under different temperature-dependent assumptions. For the existing (light blue) approach, we allow the transmission cycle to change every day based on the temperature observed that day. Our stage-specific approach means that the generation time distribution can vary based on temperatures observed for calendar times when the estimated (by Monte Carlo samples) generation events occurred.

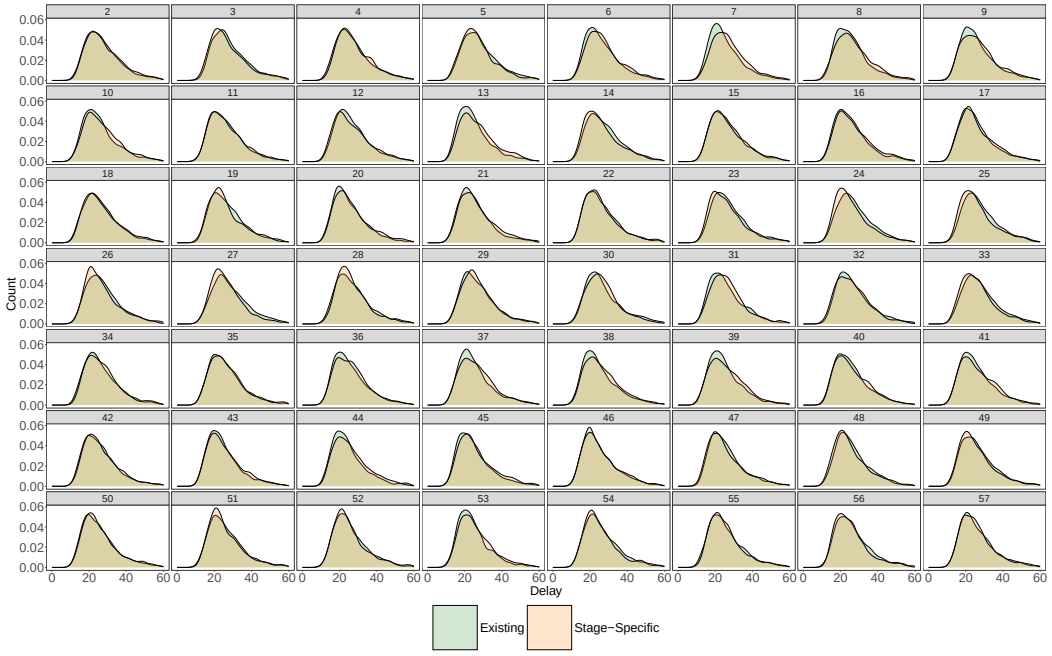


Figure SI E.4: **Time-varying generation times in São João do Tauape, Fortaleza, Brazil**. For individual calendar times, represented by numbers in facet labels, we present kernel density estimates of Monte Carlo samples from the generation time distribution under different temperature-dependent assumptions. For the existing (light blue) approach, we allow the transmission cycle to change every day based on the temperature observed that day. Our stage-specific approach means that the generation time distribution can vary based on temperatures observed for calendar times when the estimated (by Monte Carlo samples) generation events occurred.

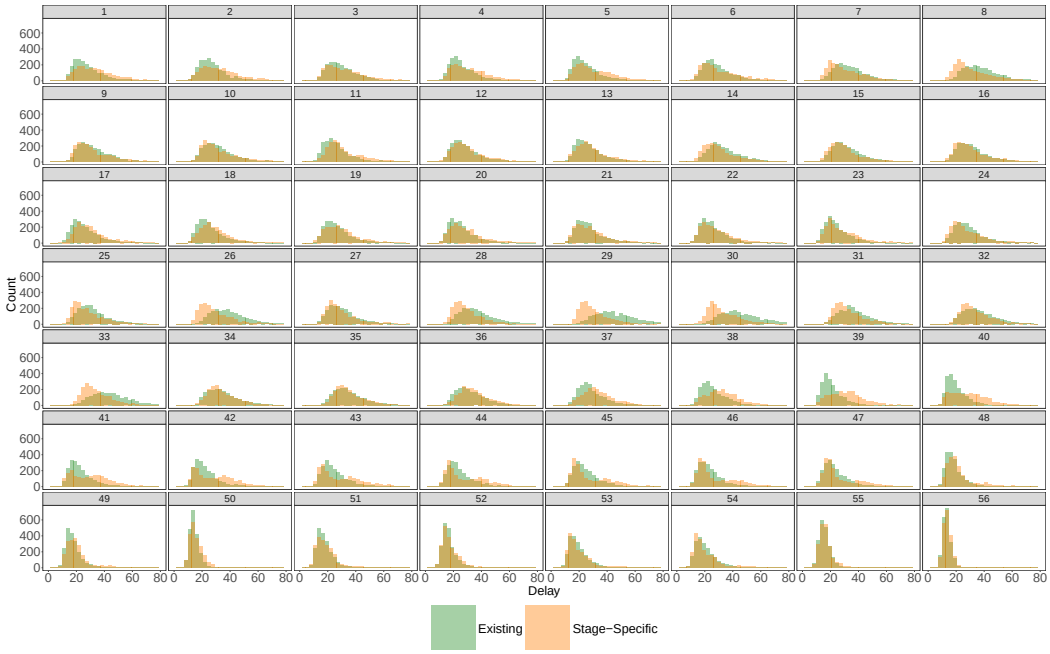


Figure SI E.5: **Time-varying generation times in our simulated environment**. For individual calendar times, represented by numbers in facet labels, we present histograms of Monte Carlo samples from the generation time distribution under different temperature-dependent assumptions. For the existing (light blue) approach, we allow the transmission cycle to change every day based on the temperature observed that day. Our stage-specific approach means that the generation time distribution can vary based on temperatures observed for calendar times when the estimated (by Monte Carlo samples) generation events occurred.

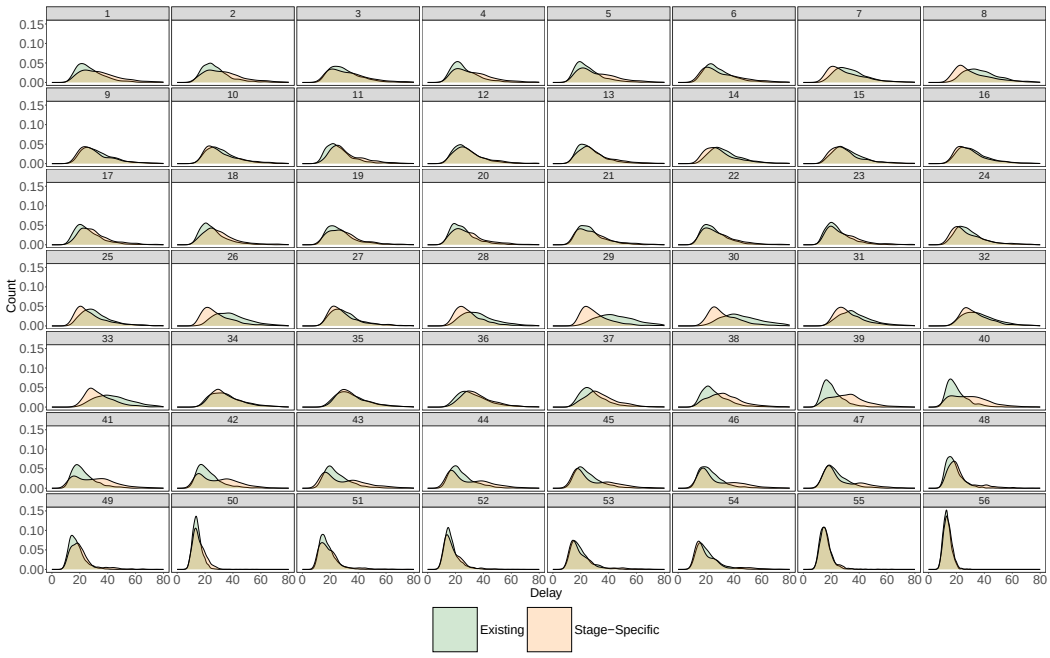


Figure SI E.6: **Time-varying generation times in our simulated environment**. For individual calendar times, represented by numbers in facet labels, we present kernel density estimates of Monte Carlo samples from the generation time distribution under different temperature-dependent assumptions. For the existing (light blue) approach, we allow the transmission cycle to change every day based on the temperature observed that day. Our stage-specific approach means that the generation time distribution can vary based on temperatures observed for calendar times when the estimated (by Monte Carlo samples) generation events occurred.

Appendix F

Appendix — Exact simulation of stochastic reaction processes with external variability

F.1 Further results for simple chemical reaction systems

F.1.1 Exponentially decreasing rates

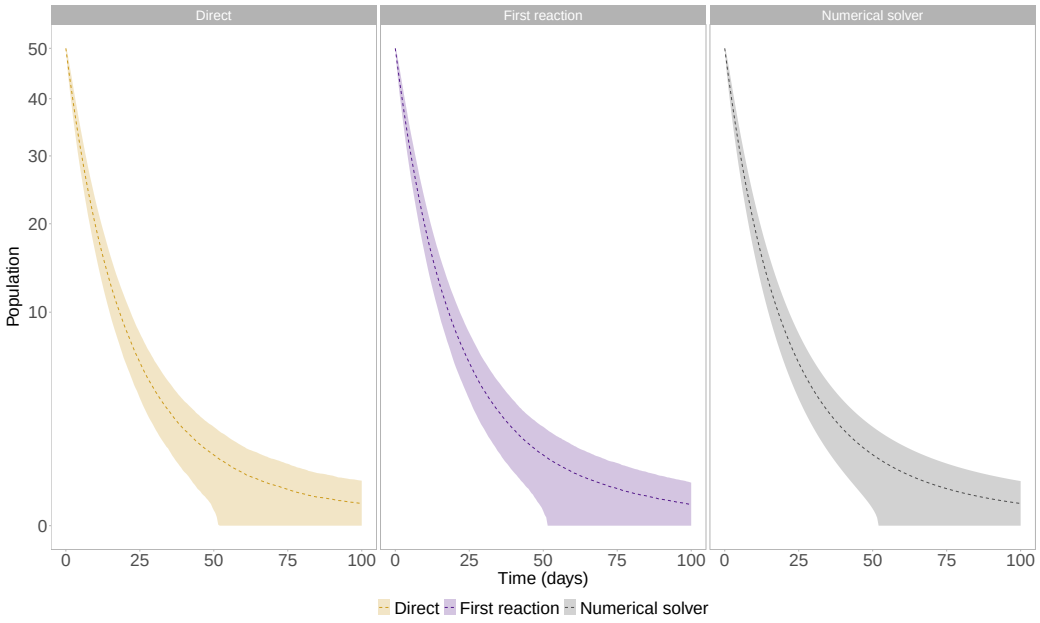


Figure SI F.1: **Variability of trajectories for system chemical reaction systems under exponentially decreasing rates:** For a chemical reaction system described by eqs (7.4.2) – (7.4.5), we illustrate further simulation results for the Gillespie direct and first reaction algorithms (Algorithms (4) and (6) respectively) and numerical ODE solver under assumptions of exponentially decreasing birth and death rates ($b(t)$ and $d(t)$ respectively). These rates are defined using eqs (7.4.8) – (7.4.9) with $c_1 = c_2 = 0.1$, $k_1 = 0.075$, and $k_2 = 0.01$. Each facet summarises trajectories from the Gillespie's direct and first reaction methods and from the numerical ODE solver. We used $\zeta = 2,500$ trajectories for the Gillespie algorithms, $n_{\max} = 400$ for the ODE solver, and summarised trajectories for all methods at interpolated time steps, from 0 to 100 at steps of size 0.5. Uncertainty intervals for the simulated trajectories of the system show the mean ± one standard deviation. See main text for more details.

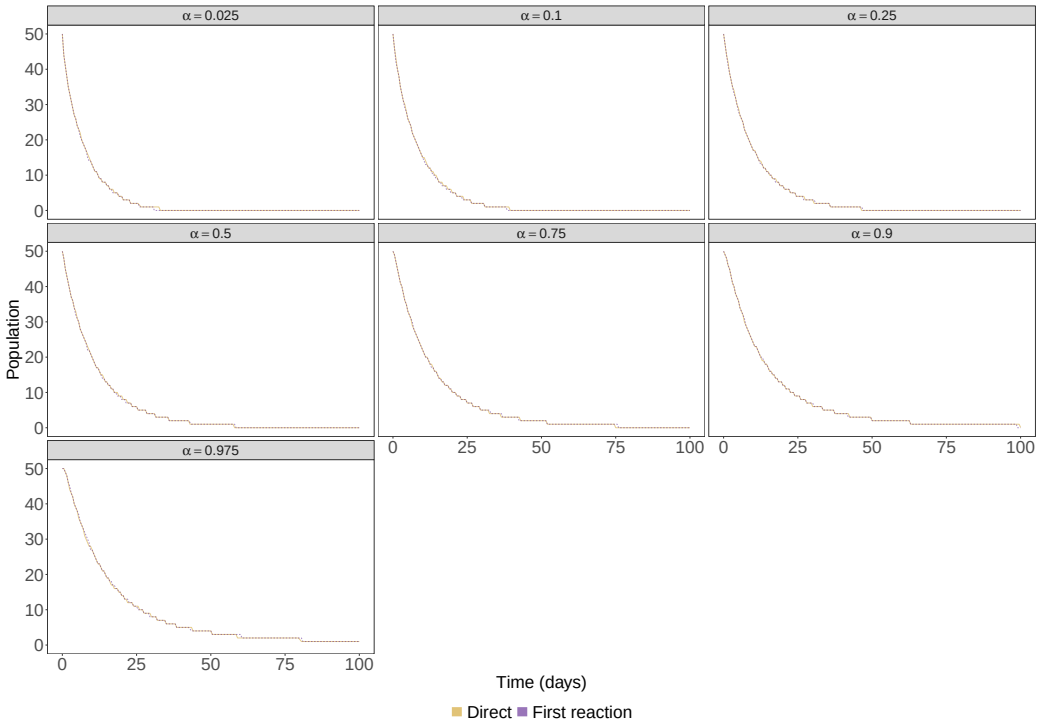


Figure SI F.2: **Distribution of trajectories for system chemical reaction systems under exponentially decreasing rates:** For a chemical reaction system described by eqs (7.4.2) – (7.4.5), we illustrate further simulation results for the Gillespie direct and first reaction algorithms (Algorithms (4) and (6) respectively) under assumptions of the exponentially decreasing birth and death rates ($b(t)$ and $d(t)$ respectively). These rates are defined using eqs (7.4.8) – (7.4.9) with $c_1 = c_2 = 0.1$, $k_1 = 0.075$, and $k_2 = 0.01$. Each facet describes a quantile level (α) of the distribution of the $\zeta = 2,500$ trajectories at each interpolated time step, from 0 to 100 at steps of size 0.5. See main text for more details.

F.1.2 Linearly increasing rates

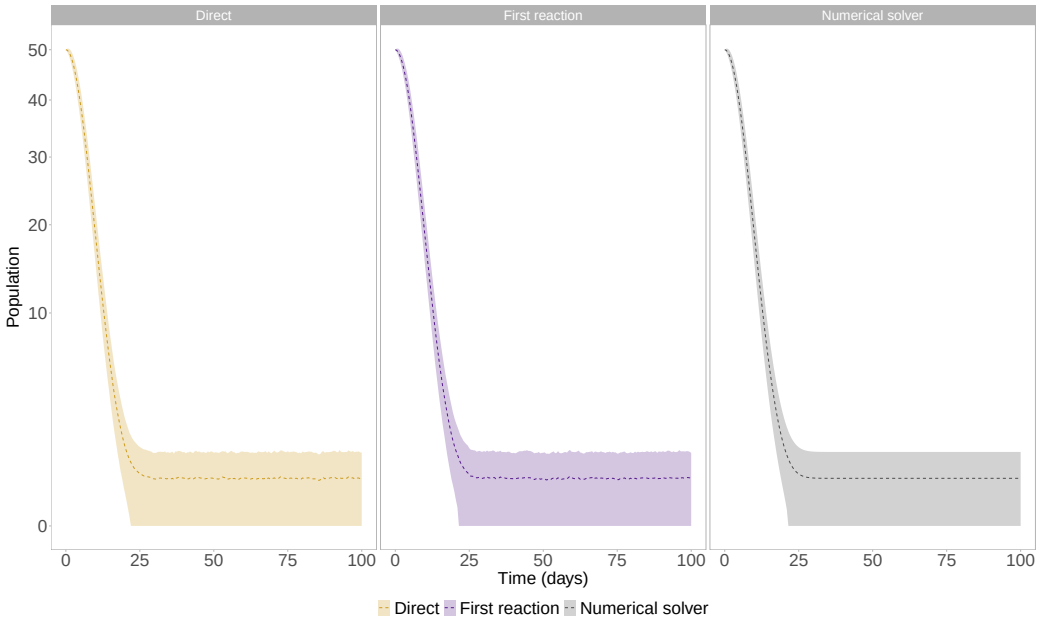


Figure SI F.3: **Variability of trajectories for system chemical reaction systems under linearly increasing rates:** For a chemical reaction system described by eqs (7.4.2) – (7.4.5), we illustrate further simulation results for the Gillespie direct and first reaction algorithms (Algorithms (4) and (6) respectively) and numerical ODE solver under assumptions of linearly increasing birth and death rates ($b(t)$ and $d(t)$ respectively). These rates are defined using eqs (7.4.12) – (7.4.13) with $c_1 = c_2 = 0.1$, $k_1 = 0.075$, and $k_2 = 0.01$. These rates are defined using eqs (7.4.12) – (7.4.13) with $c_1 = c_2 = 0.1$, $k_1 = 0.075$, and $k_2 = 0.01$. Each facet summarises trajectories from the Gillespie's direct and first reaction methods and from the numerical ODE solver. We used $\zeta = 2,500$ trajectories for the Gillespie algorithms, $n_{\max} = 400$ for the ODE solver, and summarised trajectories for all methods at interpolated time steps, from 0 to 100 at steps of size 0.5. Uncertainty intervals for the simulated trajectories of the system show the mean ± one standard deviation. See main text for more details.

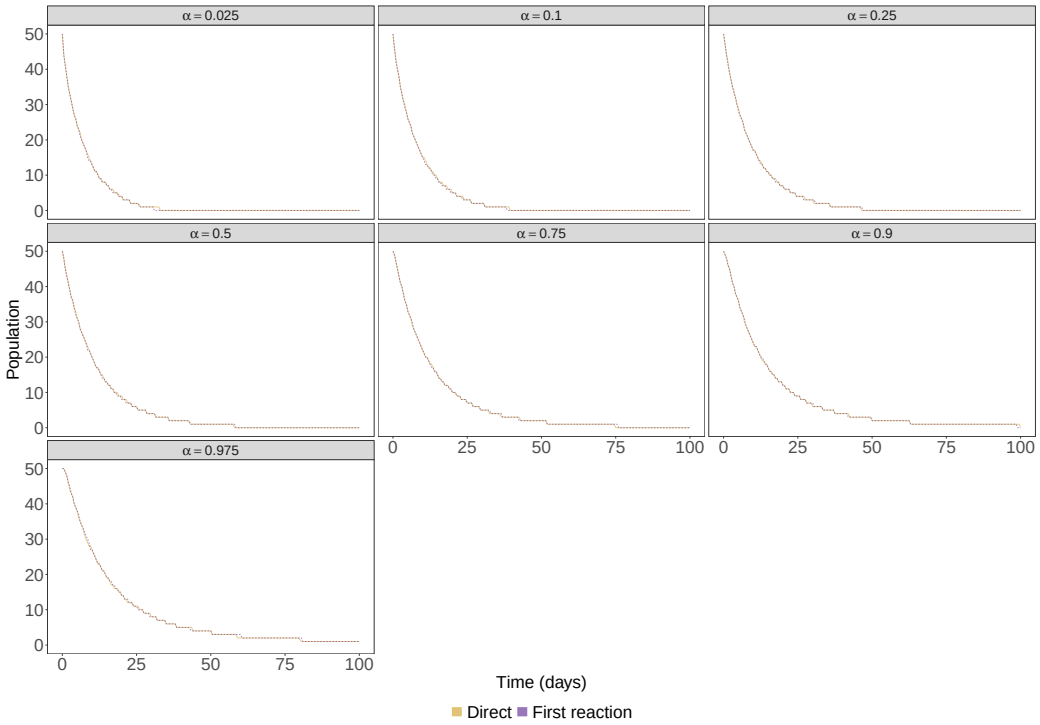


Figure SI F.4: **Distribution of trajectories for system chemical reaction systems under linearly increasing rates:** For a chemical reaction system described by eqs (7.4.2) – (7.4.5), we illustrate further simulation results for the Gillespie direct and first reaction algorithms (Algorithms (4) and (6) respectively) under assumptions of the linearly increasing birth and death rates ($b(t)$ and $d(t)$ respectively). These rates are defined using defined using eqs (7.4.12) – (7.4.13) with $l_1 = 0.01$ and $l_2 = 0.02$. Each facet describes a quantile level (α) of the distribution of the $\zeta = 2,500$ trajectories at each interpolated time step, from 0 to 100 at steps of size 0.5. See main text for more details.

F.2 Details and results of application to system of dengue transmission

F.2.1 Symbols of application

Table SI F.1: **Table of notation for individual-based model of dengue transmission:** For the system described by eqs (7.5.1) – (7.5.8), we describe all notation used.

Symbol	Definition
Φ	A state that denotes all individuals (mosquitoes and humans) no longer relevant to the process of dengue transmission.
$S^M(t, \mathbf{a}_\mathbf{S}^\mathbf{M})$	Vector consisting of the susceptible mosquitoes at time t which have ages $\mathbf{a}_\mathbf{S}^\mathbf{M}$.
$s^M(t, a_s^M)$	An individual susceptible mosquito of age a_s^M at time t.
$W(t, \mathbf{a}_\mathbf{W})$	Vector consisting of the exposed/infected (but not yet infectious) mosquitoes at time t which have ages $\mathbf{a}_\mathbf{W}$.
$w(t, a_w)$	An individual exposed/infected (but not yet infectious) mosquito of age a_w at time t.
$V(t, \mathbf{a}_\mathbf{V})$	Vector consisting of the infectious mosquitoes at time t which have ages $\mathbf{a}_\mathbf{V}$.
$v(t, a_v)$	An individual infectious mosquito of age a_v at time t.
$s^M(t, a_s^M)$	An individual susceptible mosquito of age a_s^M at time t.
$S^H(t, \mathbf{a}_\mathbf{S}^\mathbf{H})$	Vector consisting of the susceptible humans at time t which have ages $\mathbf{a}_\mathbf{S}^\mathbf{H}$.
$s^H(t, a_s^H)$	An individual susceptible human of age a_s^M at time t.
$E(t, \mathbf{a}_\mathbf{E})$	Vector consisting of the exposed/infected (but not yet infectious) humans at time t which have ages $\mathbf{a}_\mathbf{E}$.
$e(t, a_e)$	An individual exposed/infected (but not yet infectious) mosquito of age a_e at time t.
$I(t, \mathbf{a}_\mathbf{I})$	Vector consisting of the infectious humans at time t which have ages $\mathbf{a}_\mathbf{I}$.
$i(t, a_i)$	An individual infectious human of age a_i at time t.
$\mathbf{a}_\mathbf{E}$	Vector of exposed human population ages.
$\mathbf{a}_\mathbf{I}$	Vector of infectious human population ages.
$\mathbf{a}_\mathbf{S}^\mathbf{M}$	Vector of susceptible mosquito population ages.
$\mathbf{a}_\mathbf{W}$	Vector of exposed mosquito population ages.
$\mathbf{a}_\mathbf{V}$	Vector of infectious mosquito population ages.
b	Biting rate of mosquito per day.
N^H	Total count for human population.
$\lambda^M(t)$	Time-varying number of susceptible mosquitoes born per day.
μ^I	Removal rate of an individual infectious human per day.
μ^M	Death rate of an individual mosquito per day.
p_{EV}	Probability of human infection upon being bitten by an infectious mosquito.
p_{WI}	Probability of mosquito infection upon biting an infectious human.

F.2.2 Parametric assumptions

For the system described by eqs (7.5.1) – (7.5.8), we list assumptions used in our simulations:

$$\lambda^M(t) = 10 + 2 \cdot \sin\left(\frac{2\pi t}{365}\right) \tag{F.2.1}$$

$$b = 0.23 \tag{F.2.2}$$

$$\mu^M = 0.1 \tag{F.2.3}$$

$$\mu^I = 0.2 \tag{F.2.4}$$

$$p_{EV} = 0.75 \tag{F.2.5}$$

$$p_{WI} = 0.3 \tag{F.2.6}$$

$$\frac{1}{\lambda^W_{\text{Latent}}} \sim \text{Log-Normal}\left(\exp(2.9 - 0.08 \cdot 27), \frac{1}{4.9}\right), \tag{F.2.7}$$

$$\frac{1}{\lambda^E_{\text{Latent}}} \sim \text{Log-Normal}\left(\exp(0.56), \frac{1}{13.7}\right), \tag{F.2.8}$$

where we used typical values for the fixed parameters $b, \mu^M, \mu^I, p_{EV}, p_{WI}$, the temperature-dependent assumptions of Siraj et al.[156] for the extrinsic latent period $\frac{1}{\lambda^W_{\text{Latent}}}$, alongside a constant temperature of 27 degrees celsius, and the time-invariant assumptions of Siraj et al.[156] for the intrinsic latent period $\frac{1}{\lambda^E_{\text{Latent}}}$. Log-Normal distributions are of the form Log-Normal (μ, σ) with μ and σ equal to the mean and standard deviation on a logarithmic scale. Our objective was simply to illustrate simulation for an elaborate system, but real-world applications should consider the sensitivity to these assumptions.

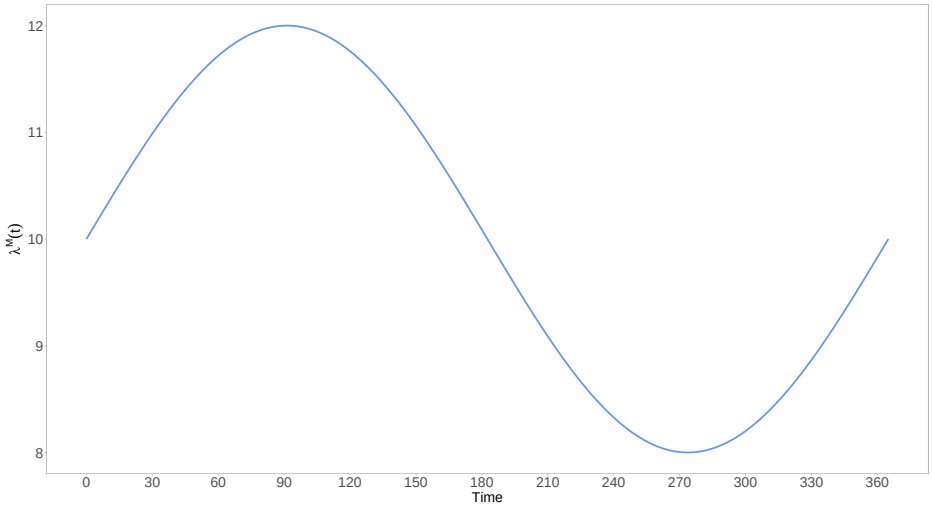


Figure SI F.5: **Time-varying event propensity for susceptible mosquito births:** The figure visualises our assumption for $\lambda^M(t)$ – the expected number of susceptible mosquitoes born per unit of time across a given year, otherwise known as the event propensity. The full system is described in eqs (7.5.1) – (7.5.8), while a table of notation is provided in Table SI F.1.

Appendix G

Appendix — From metric to action: An evaluation framework to translate infectious disease forecasts into policy decisions

G.1 Extended table of concepts

Table SI G.1: **Key concepts for our framework of probabilistic infectious disease forecasting:** We define here important concepts that are discussed throughout the text, and are integral to our proposed framework for forecasting infectious diseases. Given the wealth of metrics and tools above, the choice of metric/tool may not be immediately obvious, and may be both user- and event-specific, thus starting with the user defining their policy question(s), outcome to measure, and corresponding preferences. See Figure 8.1 for further details.

Concept	Meaning
Action set	The set of available decisions, strategies, or interventions available to decision-makers before or during an infectious disease outbreak.
Bias	The propensity of a model's forecasts to overpredict/underpredict.
Brier Score (BS)	A strictly proper scoring rule defined as mean squared error for probabilistic forecast of a binary outcome [637].
Calibration	In the context of forecast evaluation, calibration is equivalent to reliability and is the tendency for forecasted probabilities to align with observed frequencies for outcomes. There exists both conditional and unconditional calibration [648,673], while beyond forecast evaluation (and the current paper), the post-processing procedure of re-calibrating probabilistic forecasts can also be termed calibration.
Central prediction interval (PI) coverage	The percentage of observations that lie within the claimed central PIs. This is a measure of a forecasting model's calibration.
Classifiers	A probabilistic classifier produces a probability forecast for a binary outcome.
Consistent	A scoring function is consistent for a given functional $T(P)$ (e.g., quantile level) if the score is minimised when the forecast aligns with $T(P)$ consistent.
Continuous Ranked Probability Score (CRPS)	A proper scoring rule that is strict when the first moment of the forecast distribution is finite, and generalises mean absolute error for probabilistic forecasts of continuous outcomes [470].
Cost-Loss (C/L) ratio	In the C/L ratio decision model for binary outcome events, the C/L ratio is the cost (C) of preparation for an adverse event, divided by the avoidable loss (L) if the event occurs without any preparation [161,640]. In this model, the optimal decision is to act (i.e., prepare) if the forecasted probability exceeds this *decision threshold* C/L. Knowing plausible values (or a range) for the ratio (for a given binary event) is sufficient for studying forecast model results, and thus for metric-based decision-making.
Diagonal Score (DS)	For a binary outcome, the DS is a proper scoring rule that measures performance of probabilistic forecasts while holding fixed the relationship between a user-defined probability level (C/L) and an event probability [653]. This evaluation focuses on the diagonal of the forecast skill card, and more practically, can be perceived as a metric for vulnerable users, as if the cost of actions remains fixed across event thresholds, then low-probability events are associated with large losses.
Discrimination (DSC)	A non-negative component of the decomposition for the mean value for any score, and measures the ability of a model's forecasts to distinguish events versus non-events [641]. Higher values indicate superior discrimination ability.
Event threshold	A cut-off value used for defining whether a binary outcome occurs or not.
Forecast/predictive distribution	A probabilistic forecast for the future value of a given quantity, often represented by a parametric distribution, samples from the distribution, and/or quantiles of the distribution.
Forecast horizon	The lead time at which forecasts are made.
Forecast quality	Measure of the alignment of a model's forecasted values with observed values.
Forecast value	Measure of the ability of a model's forecasted values to inform decision-making.
Horizon	The lead time at which a forecast for a future value is produced.
Miscalibration (MSC)	A non-negative component of the decomposition for the mean value for any proper scoring rule, and measures the calibration of a model's forecasts [641]. Lower values indicate superior calibration.
Permutation entropy (PE)	Measure of predictability which work by studying temporal dependencies in both linear and nonlinear time series. PE is a model-free measure that uses Shannon's Entropy to study the frequency of distinct patterns in time series data.
Predictability	The inherent randomness of a system which impacts the extent to which future values of a system can be reliably/correctly forecasted.
Proper	A scoring rule is proper if the expected score is minimised by using the true probability distribution.
Quantile Score (QS)	A measure of forecast performance for a specific user-defined probability/quantile level. QS is also known as pinball loss or the asymmetric piecewise linear scoring function.
Rolling/trailing window	A fixed length of time, say L, which means that only the past L time series observations are included in analysis for either a given region's time series or across all regions.
Scoring function	A function that assigns a score to a given functional (e.g., quantile level). This is not necessarily proper, and can asymmetrically penalise probabilistic forecasts (e.g., QS).
Scoring rule	A scoring rule (e.g., CRPS) is a specific type of scoring function (see above) that assigns a score to a probabilistic forecast, and lower values for a scoring rule imply better predictive performance.
Skill	Performance of a forecasting model relative to a baseline or reference model. Often computed using a scoring rule (e.g., CRPS), the resulting skill score can measure relative model performance across space and time.

G.2 Relative Economic Value (REV)

The REV curve for a model is based on the C/L model presented in Section 8.2.3. We present the C/L matrix again to aid interpretation:

	Event occurs	**Event does not occur**
Action taken	C	C
Action not taken	L	0

Suppose $\theta \in \mathbb{R}$ is a fixed event threshold (e.g., number of cases). Binary events ($y = 1$ if $> \theta$) are predicted to occur if the model forecasts the event with probability greater than or equal to the user-defined C/L ratio α. This is equivalent to a forecast at quantile level 1-α being greater than or equal to θ.

Then, the expected *expense* of the model (E_{model}) is derived from the confusion matrix for the model's classifications, incurring costs C for observations where the model correctly predicted an event that occurs (a true positive rate, TP) or where the model incorrectly predicted an event that does not occur (a false positive rate, FP). On the other hand, events where the model failed to predict the event (false negative rate, FN) will incur the preventable loss L, while a true negative will incur no cost or loss. Mathematically, we write this as:

$$E_{\text{model}} = (\text{TP} + \text{FP}) \cdot C + \text{FN} \cdot L, \tag{G.2.1}$$

while a perfect (or oracle) probability model incurs expense only when the model predicts the event (which always coincides when the event occurs

$$E_{\text{perfect}} = \text{TP} \cdot C, \tag{G.2.2}$$

while a baseline (or reference) model is derived analogously to the studied model – using the baseline model's confusion matrix

$$E_{\text{baseline}} = (\text{TP}_{\text{baseline}} + \text{FP}_{\text{baseline}}) \cdot C + \text{FN}_{\text{baseline}} \cdot L. \tag{G.2.3}$$

Then, REV can be defined as:

$$\text{REV}(\alpha = C/L, \theta) := \frac{E_{\text{baseline}} - E_{\text{model}}}{E_{\text{baseline}} - E_{\text{perfect}}}. \tag{G.2.4}$$

For calculation, we see that REV is:

$$\text{REV}(\alpha = C/L, \theta) := \frac{(\text{TP}_{\text{baseline}} + \text{FP}_{\text{baseline}}) \cdot C + \text{FN}_{\text{baseline}} \cdot L - (\text{TP} + \text{FP}) \cdot C + \text{FN} \cdot L}{(\text{TP}_{\text{baseline}} + \text{FP}_{\text{baseline}}) \cdot C + \text{FN}_{\text{baseline}} \cdot L - \text{TP} \cdot C}, \tag{G.2.5}$$

where dividing through by loss L, allows us to compute REV when we only know the C/L ratio α.

G.3 EPIFORGE 2020 guidelines

Table SI G.2: **EPIFORGE 2020 guidelines:** We present the pages of our manuscript [52] which correspond to the checklist items of the EPIFORGE 2020 guidelines for forecast reporting [564].

Section of Manuscript	#	Checklist Item	Reported on Page
Title/Abstract	1	Describe the study as forecast or prediction research in at least the title or abstract.	1
Introduction	2	Define the purpose of study and forecasting targets.	1 and 14
Methods	3	Fully document the methods.	2-14
Methods	4	Identify whether the forecast was performed prospectively, in real time, and/or retrospectively.	14
Methods	5	Explicitly describe the origin of input source data, with references.	14
Methods	6	Provide source data with publication, or document reasons as to why this was not possible.	14
Methods	7	Describe input data processing procedures in detail.	14
Methods	8	State and describe the model type, and document model assumptions, including references.	14
Methods	9	Make the model code available or document the reasons why this is not possible.	14
Methods	10	Describe the model validation and justify the approach.	2-14
Methods	11	Describe the forecast accuracy evaluation method used, with justification.	2-14
Methods	12	Where possible, compare results to a benchmark or other comparator model, with justification of comparator choice.	14
Methods	13	Describe the forecast horizon, with justification of its length.	14
Results	14	Present and explain uncertainty of forecasting results.	15-19
Results	15	Briefly summarize the results in nontechnical terms, including a nontechnical interpretation of forecast uncertainty.	15-19
Results	16	If results are published as a data object, encourage a time-stamped version number.	14
Discussion	17	Describe the weaknesses of the forecast, including weaknesses specific to data quality and methods.	22
Discussion	18	If the forecast research is applicable to a specific epidemic, comment on its potential implications and impact for public health action and decision-making.	15-19, 22
Discussion	19	If the forecast research is applicable to a specific epidemic, comment on how generalizable it may be across populations.	22

n/a= Not Applicable

G.4 COVID-19 Forecast Hub Application

The COVID-19 Forecast Hub was the host for a collaborative forecasting effort in the United States[173]. Teams were invited to submit probabilistic forecasts for weekly incident COVID-19 cases and deaths at forecast horizons of one to four weeks. We focused on the forecasting of COVID-19 cases in this work, where teams submitted forecasts using seven quantile levels (0.025, 0.1, 0.25, 0.5, 0.75, 0.9, 0.975). We used the forecast data provided by[170].

For forecast evaluation in this work, we used the scoringutils, scoringrules, and triptych packages[560,663,766,767] in R version 4.4.3[584].

G.5 Additional figures

G.5.1 Weighted Interval Score (WIS) across space and time



Figure SI G.1: **Overall forecast quality across time and space:** Results for application of the relative Weighted Interval Score (rWIS) to forecasts of incident weekly COVID-19 cases at one- and four-week horizons. rWIS is a scaled measure of relative performance (relative to the baseline model) that accounts for potentially different numbers of forecasts for different locations and time points. Each plot panel visualises rWIS **A)** across all locations, **B)** for specific locations, and **C)** in trailing windows across all locations.

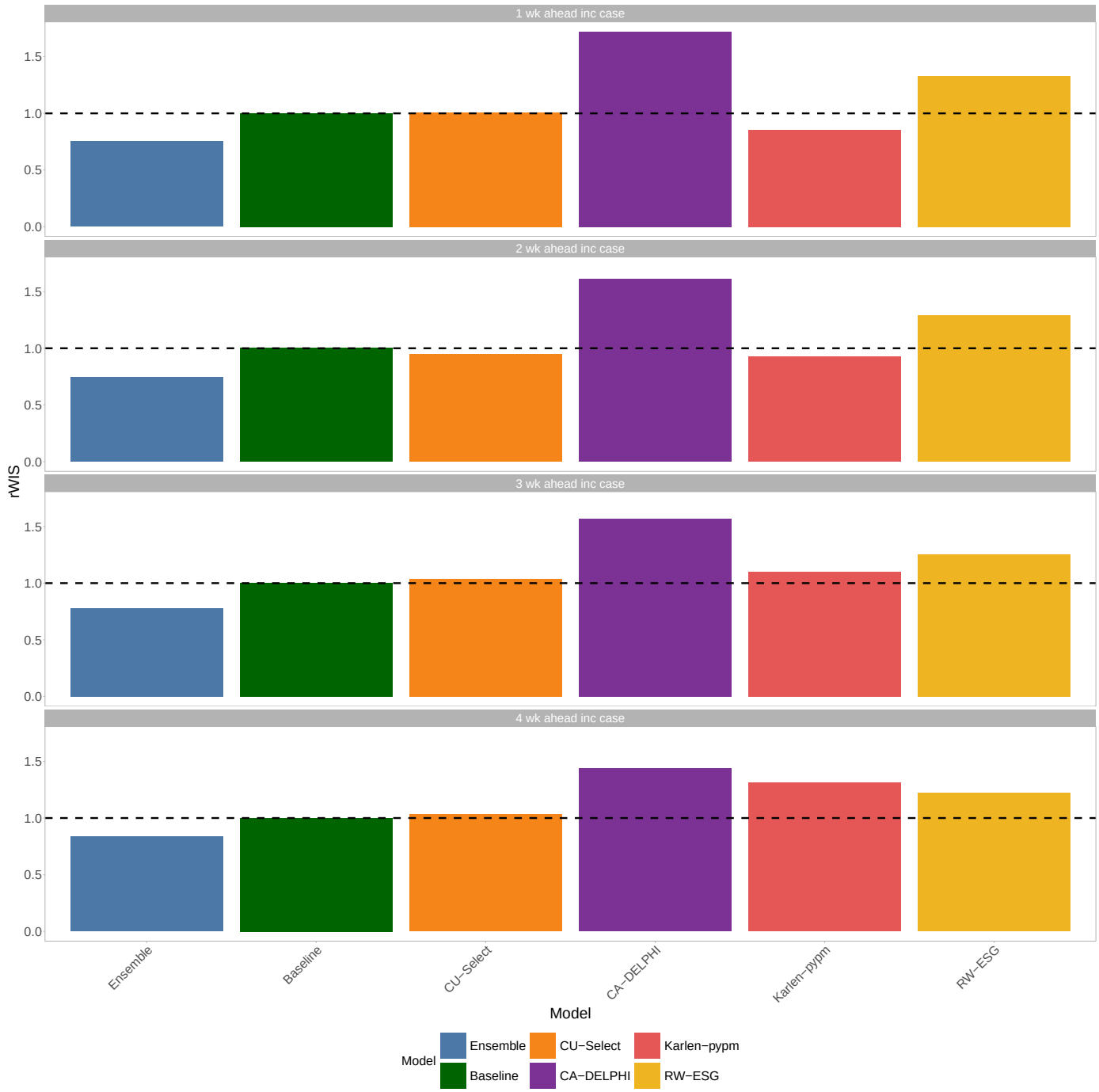


Figure SI G.2: **Relative Weighted Interval Score (WIS) by different forecast horizons across all locations:** Results for application of the Weighted Interval Score (WIS) to forecasts of incident weekly COVID-19 cases at one- to four-week horizons.

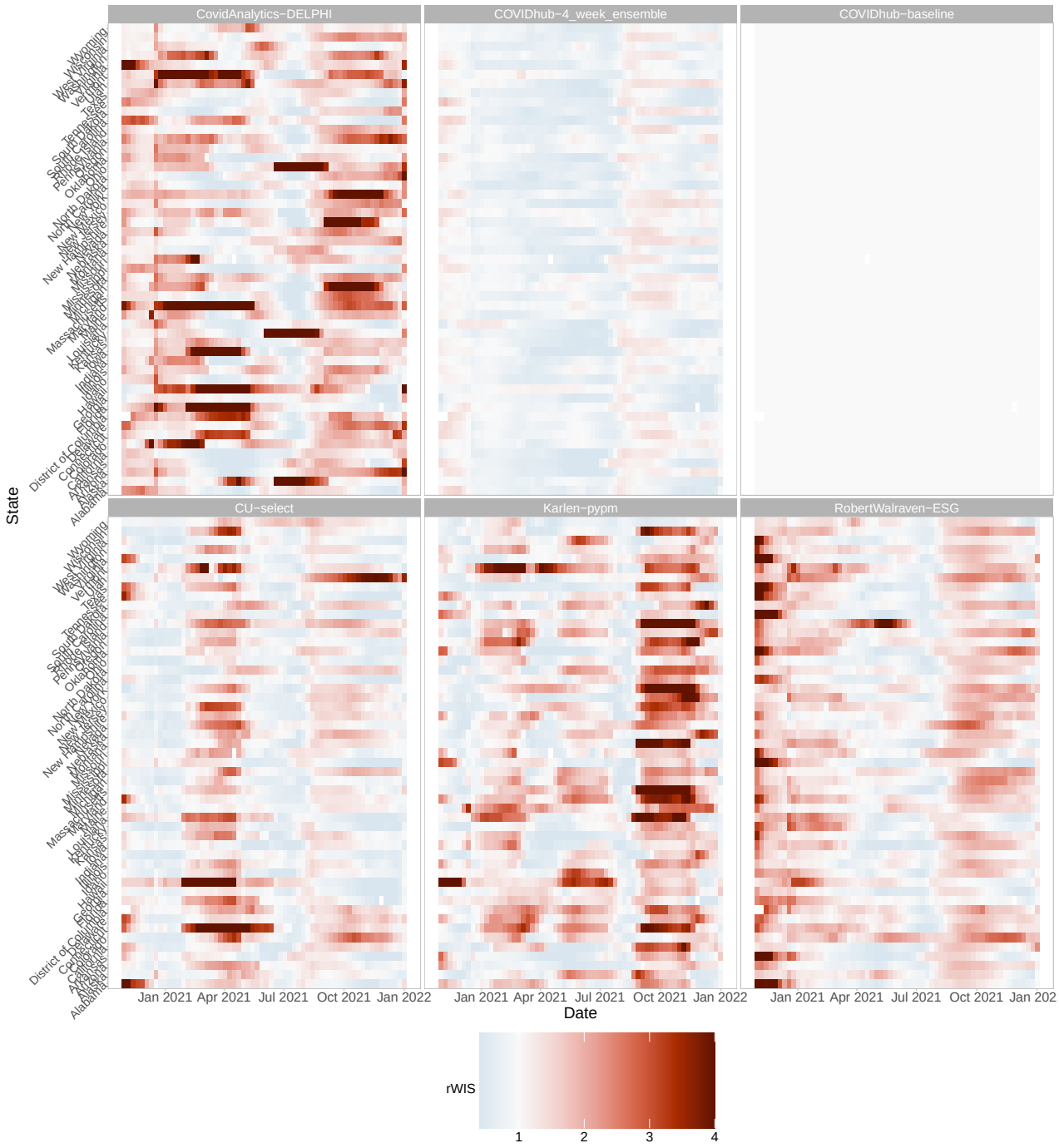


Figure SI G.3: **Relative Weighted Interval Score (rWIS) for one-week-ahead forecasts by model and location in 12-week trailing windows:** Scaled relative Weighted Interval Score[58] for different models is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rWIS relative to the baseline model, and rWIS > 1 indicates greater forecast value for the uniform family of decision-makers in the given location and trailing window.

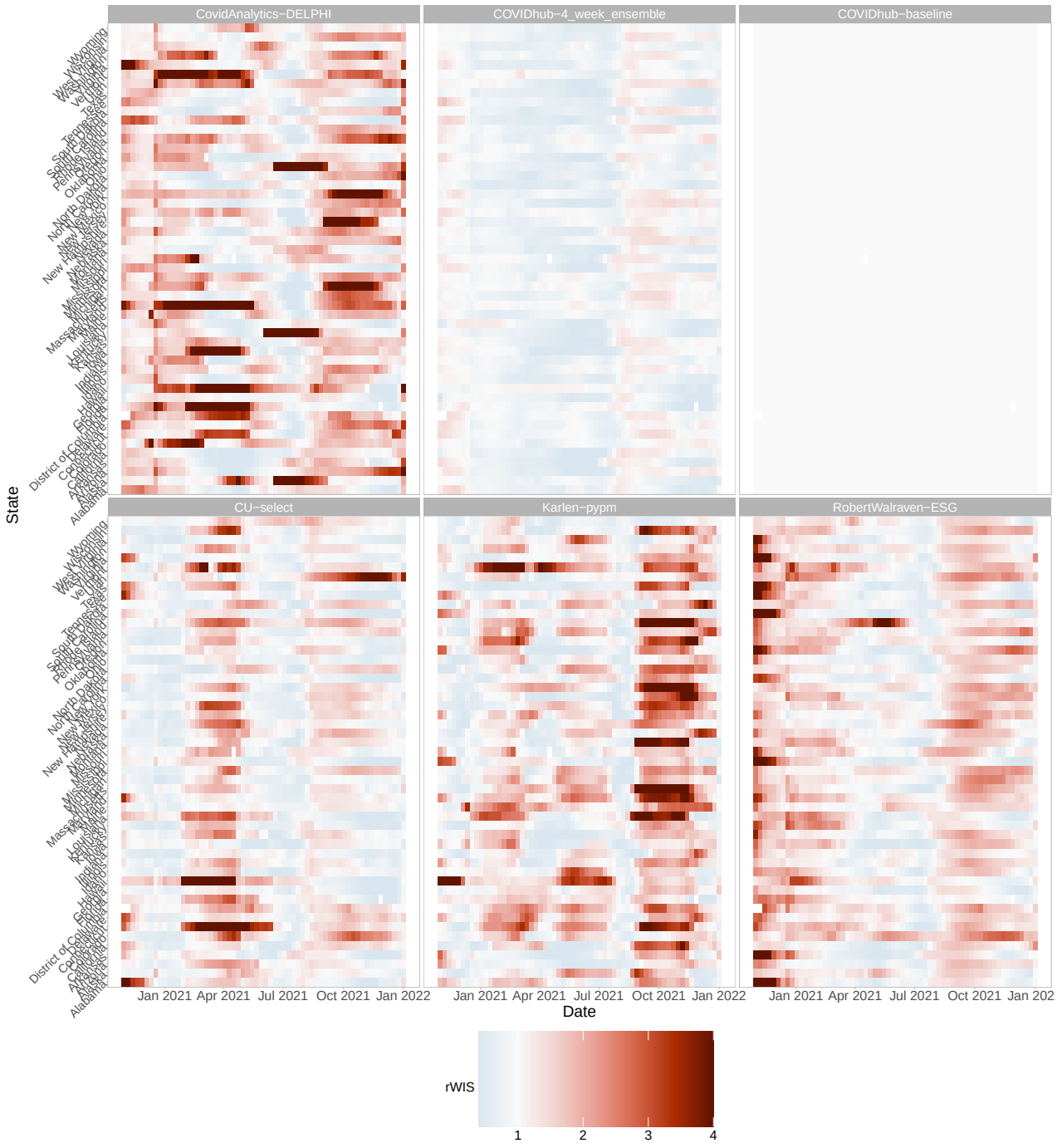


Figure SI G.4: **Relative Weighted Interval Score (rWIS) for two-weeks-ahead forecasts by model and location in 12-week trailing windows:** Scaled relative Weighted Interval Score[58] for different models is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rWIS relative to the baseline model, and rWIS > 1 indicates greater forecast value for the uniform family of decision-makers in the given location and trailing window.

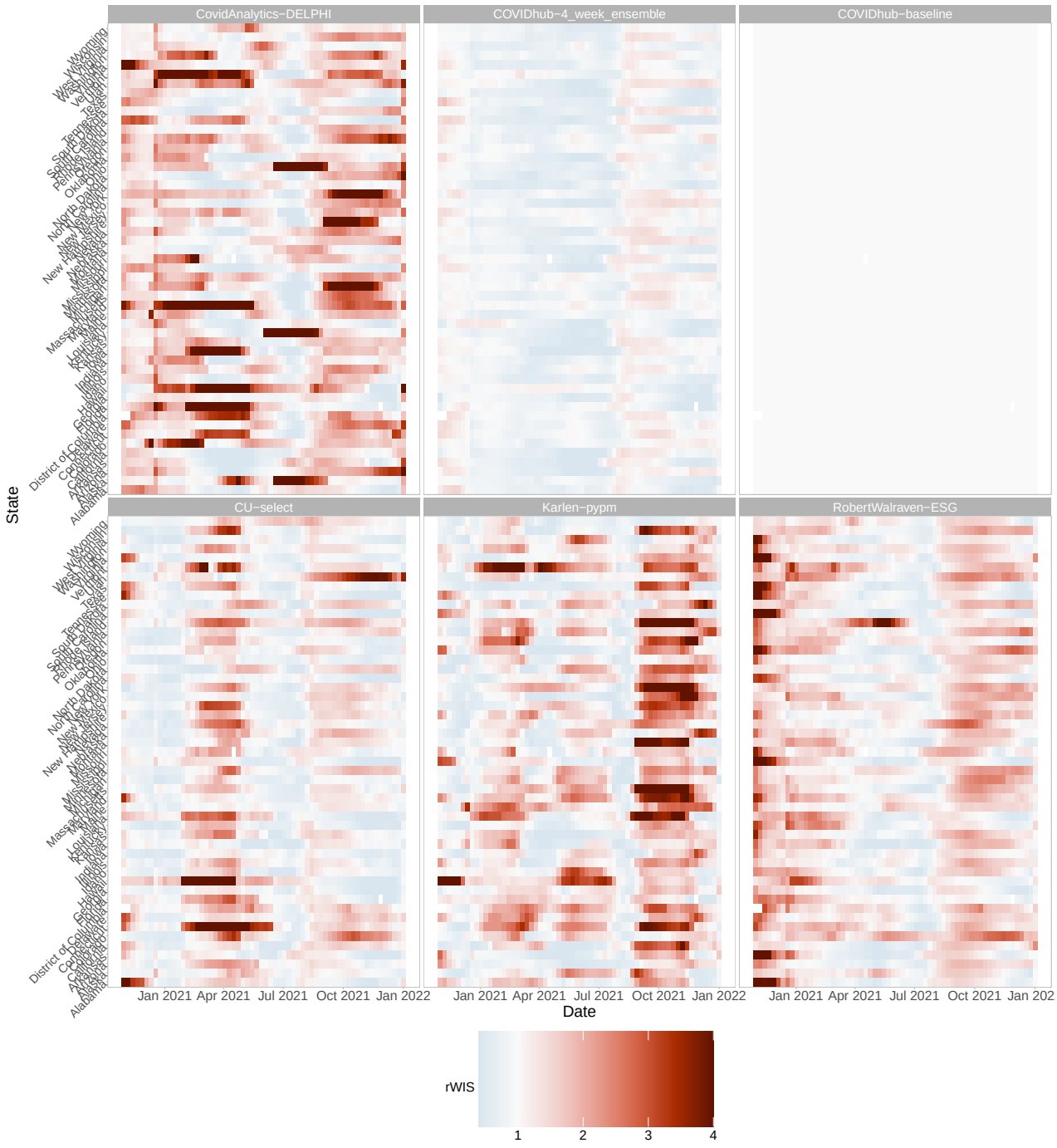


Figure SI G.5: **Relative Weighted Interval Score (rWIS) for three-weeks-ahead forecasts by model and location in 12-week trailing windows:** Scaled relative Weighted Interval Score[58] for different models is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rWIS relative to the baseline model, and rWIS > 1 indicates greater forecast value for the uniform family of decision-makers in the given location and trailing window.

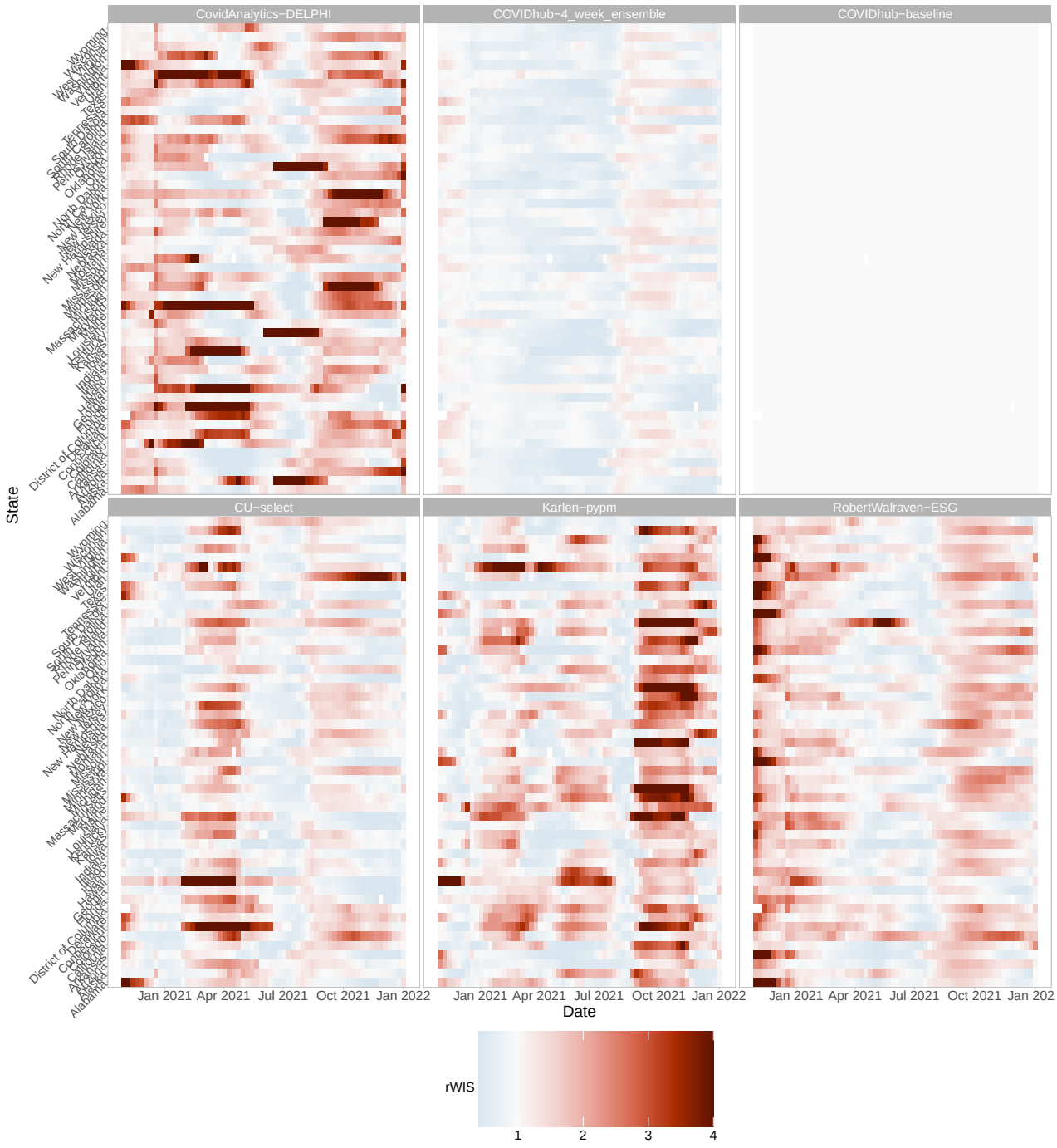


Figure SI G.6: **Relative Weighted Interval Score (rWIS) for four-weeks-ahead forecasts by model and location in 12-week trailing windows:** Scaled relative Weighted Interval Score[58] for different models is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rWIS relative to the baseline model, and rWIS > 1 indicates greater forecast value for the uniform family of decision-makers in the given location and trailing window.

G.5.2 Relative Economic Value (REV) – Differing events and reference models

Upper 10% of all events

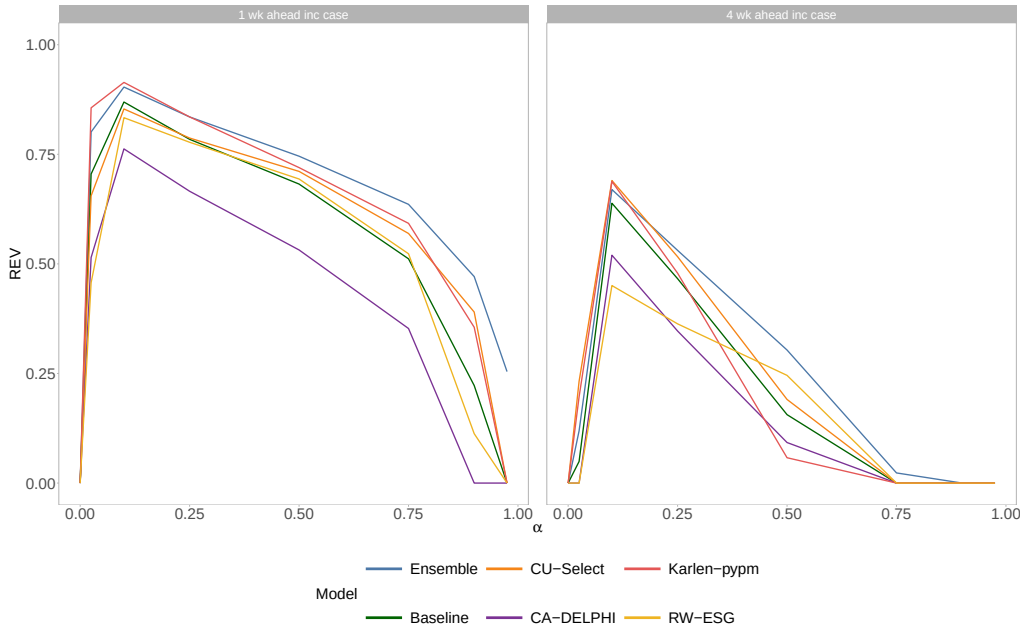


Figure SI G.7: **Relative Economic Value (REV) for upper 10% of *all* observed cases:** REV for the different forecasting models, compared to a forecasting model which forecasts the binary event with probability of 10% (i.e., analogous to the climatology model used in weather forecasting).

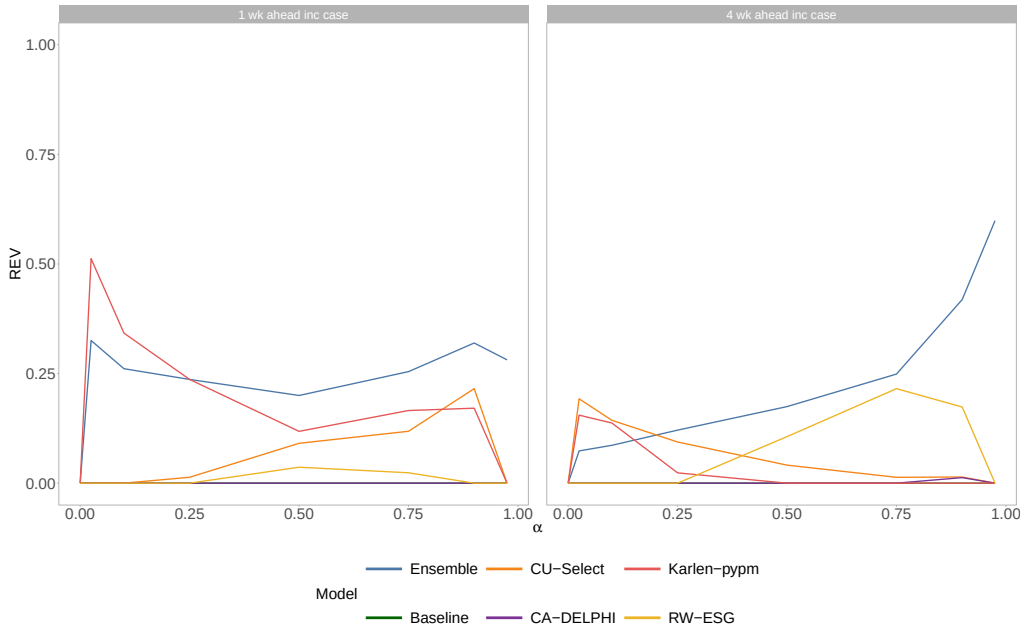


Figure SI G.8: **Relative Economic Value (REV) for upper 10% of *all* observed incident cases:** REV for the different forecasting models, compared to the COVID-19 Forecast Hub baseline forecasting model which forecasts the most recent observed incident cases as its best (i.e., median) estimate of the future.

Upper 10% of each state's events

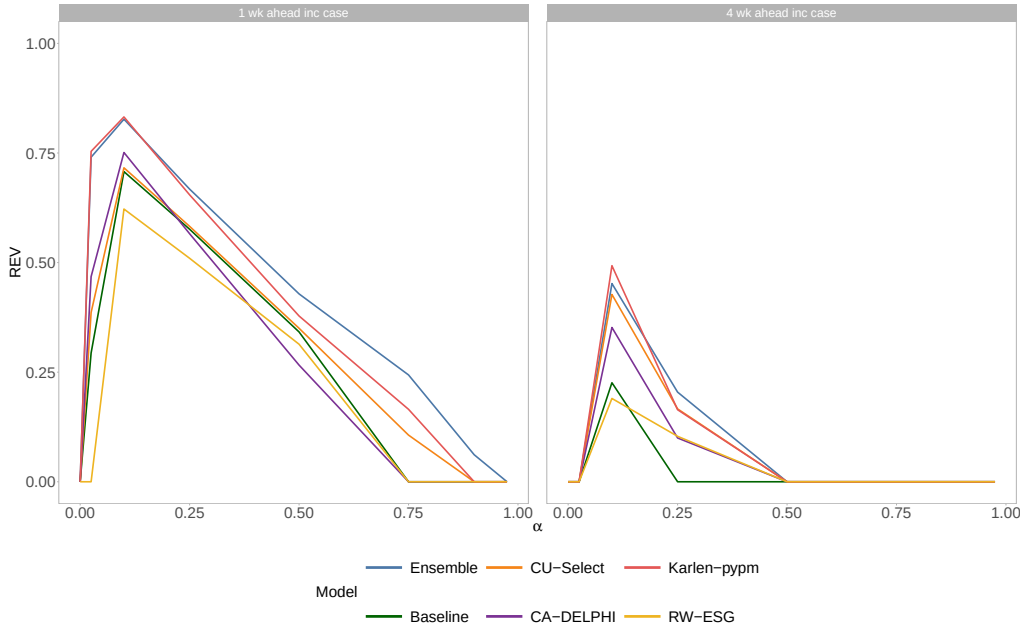


Figure SI G.9: **Relative Economic Value (REV) for upper 10% of each state's observed incident cases:** REV for the different forecasting models, compared to a forecasting model which forecasts the binary event with probability of 10% (i.e., analogous to the climatology model used in weather forecasting). This differs to Figures SI G.7 – SI G.8 by focusing on value for the upper 10% of events (i.e., cases) observed in each state.

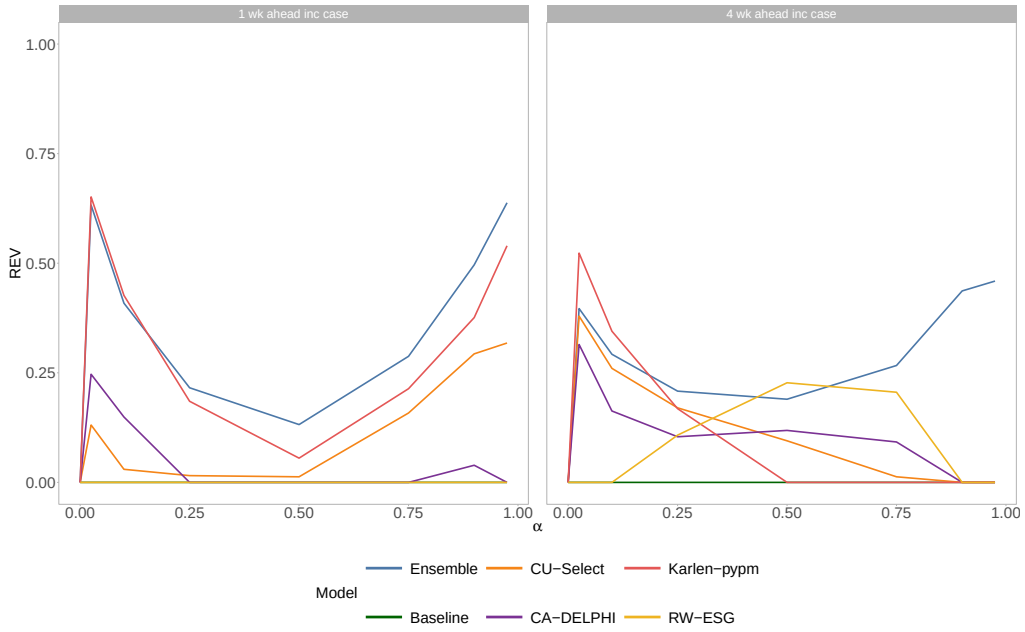


Figure SI G.10: **Relative Economic Value (REV) for upper 10% of each location's observed incident cases:** REV for the different forecasting models, compared to the COVID-19 Forecast Hub baseline forecasting model which forecasts the most recent observed incident cases as its best (i.e., median) estimate of the future.

Upper 15% of all events

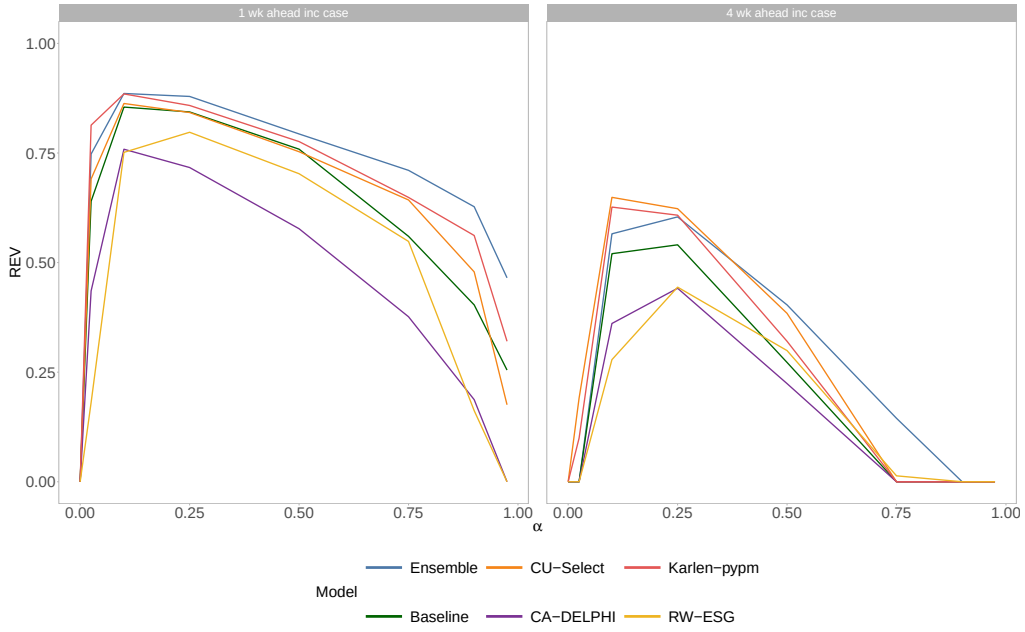


Figure SI G.11: **Relative Economic Value (REV) for upper 15% of all observed incident cases:** REV for the different forecasting models, compared to a forecasting model which forecasts the binary event with probability of 15% (i.e., analogous to the climatology model used in weather forecasting). This plot has featured in Figure 8.3 C.3.

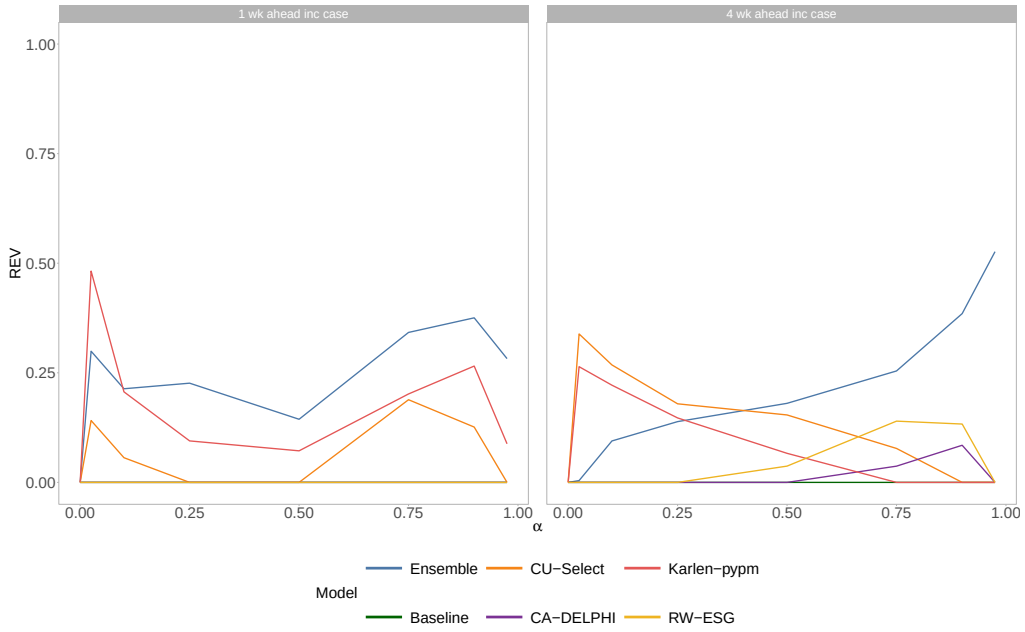


Figure SI G.12: **Relative Economic Value (REV) for upper 15% of all observed incident cases:** REV for the different forecasting models, compared to the COVID-19 Forecast Hub baseline forecasting model which forecasts the most recent observed incident cases as its best (i.e., median) estimate of the future.

Upper 15% of each state's events

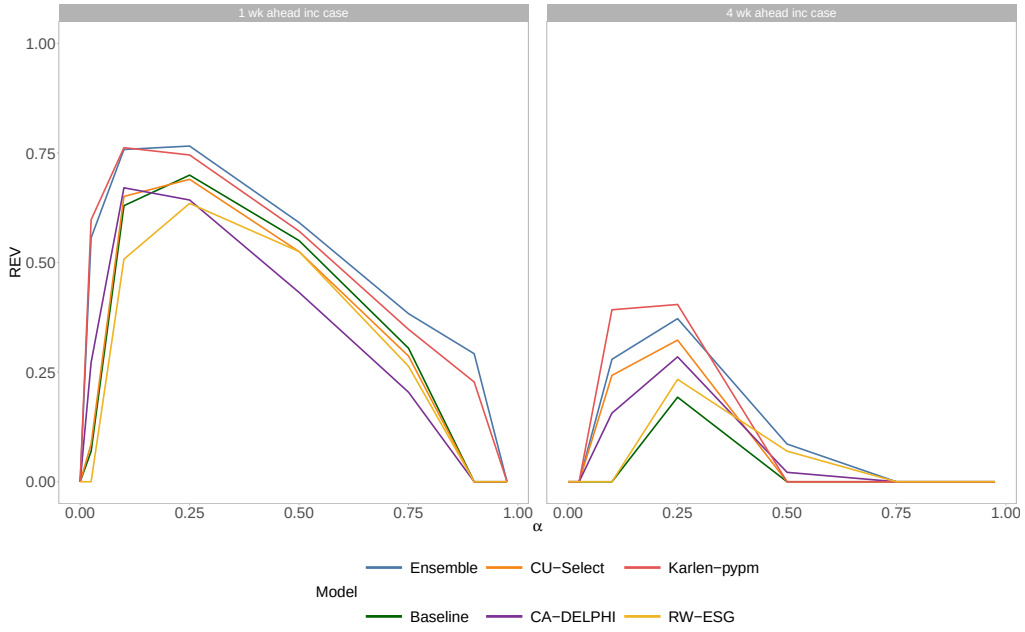


Figure SI G.13: **Relative Economic Value (REV) for upper 15% of each location's observed incident cases:** REV for the different forecasting models, compared to a forecasting model which forecasts the binary event with probability of 15% (i.e., analogous to the climatology model used in weather forecasting). This differs to Figures SI G.11 – SI G.12 by focusing on value for the upper 15% of events (i.e., cases) observed within each state.

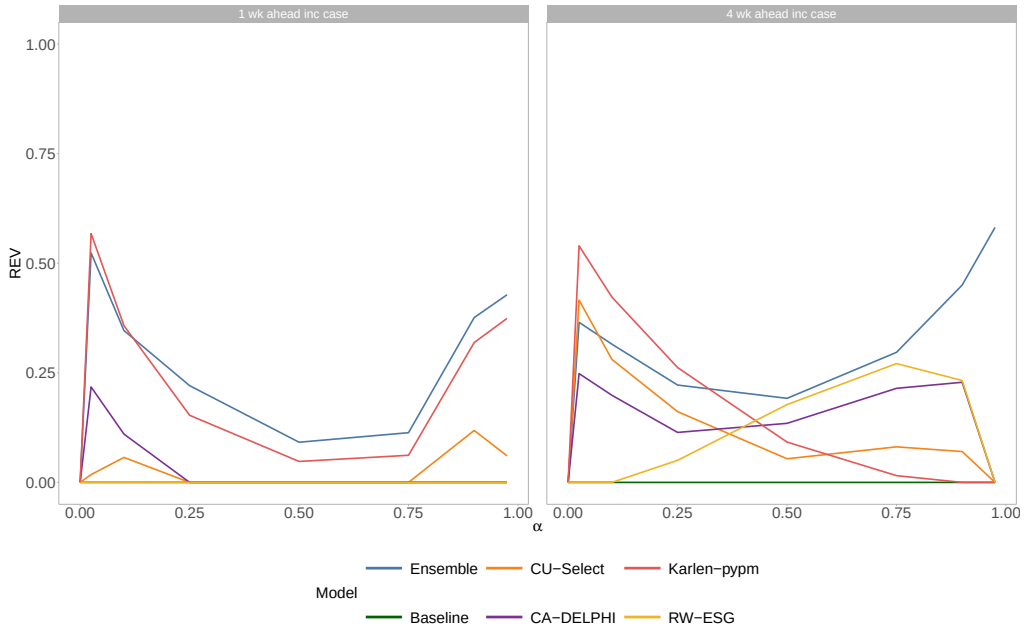


Figure SI G.14: **Relative Economic Value (REV) for upper 15% of each location's observed incident cases:** REV for the different forecasting models, compared to the COVID-19 Forecast Hub baseline forecasting model which forecasts the most recent observed incident cases as its best (i.e., median) estimate of the future.

G.5.3 Relative pinball loss (rPL) across space and time

Median rPL for models across locations over time for risk-intolerant and risk-tolerant decision-makers

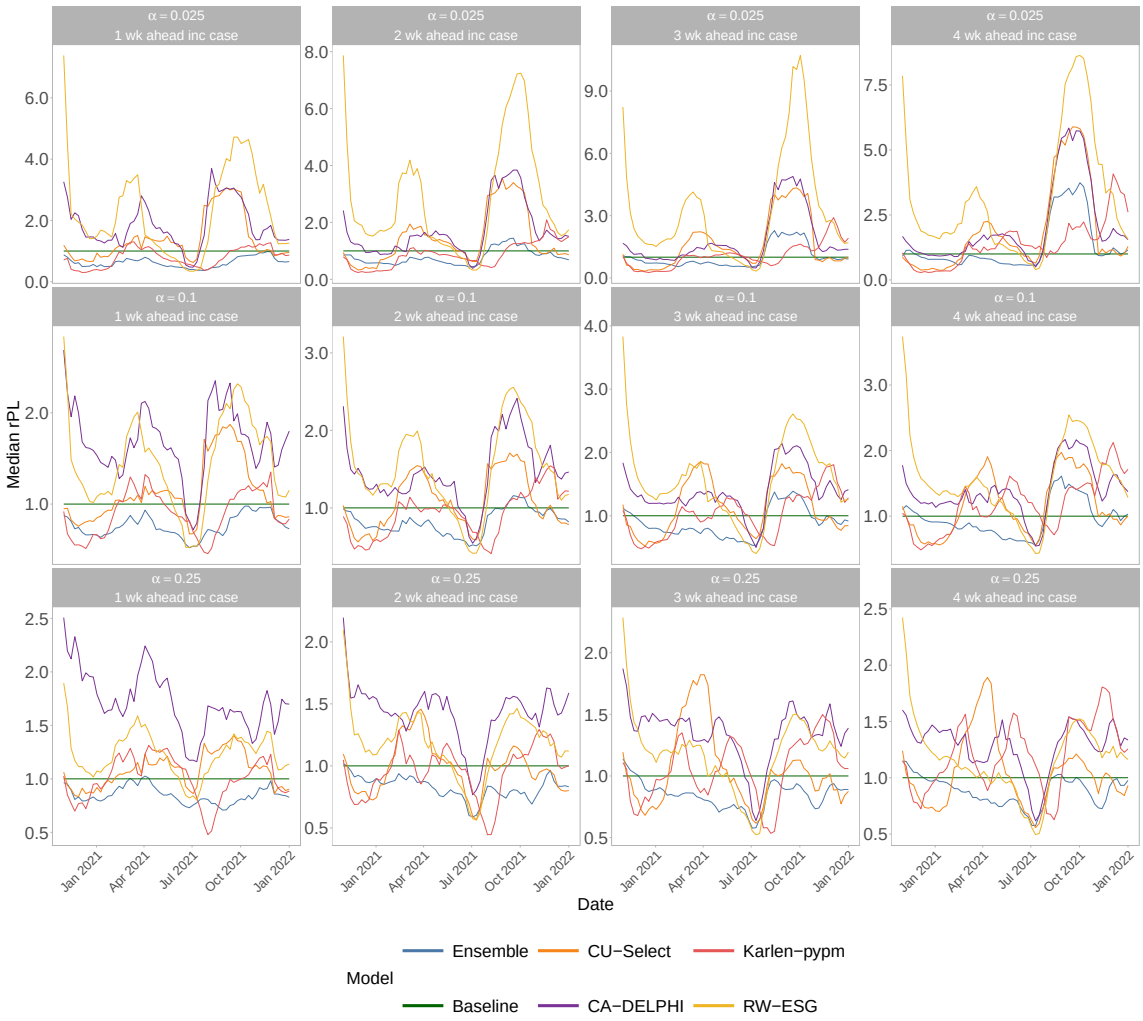


Figure SI G.15: **Median relative pinball loss (rPL) for each model's forecasts across individual locations from 12-week trailing windows:** Focusing on decision-makers with low C/L ratios (i.e., cost of action is inexpensive relative to harm), median rPL across all states (y-axis) is visualised for each model and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the trailing window.

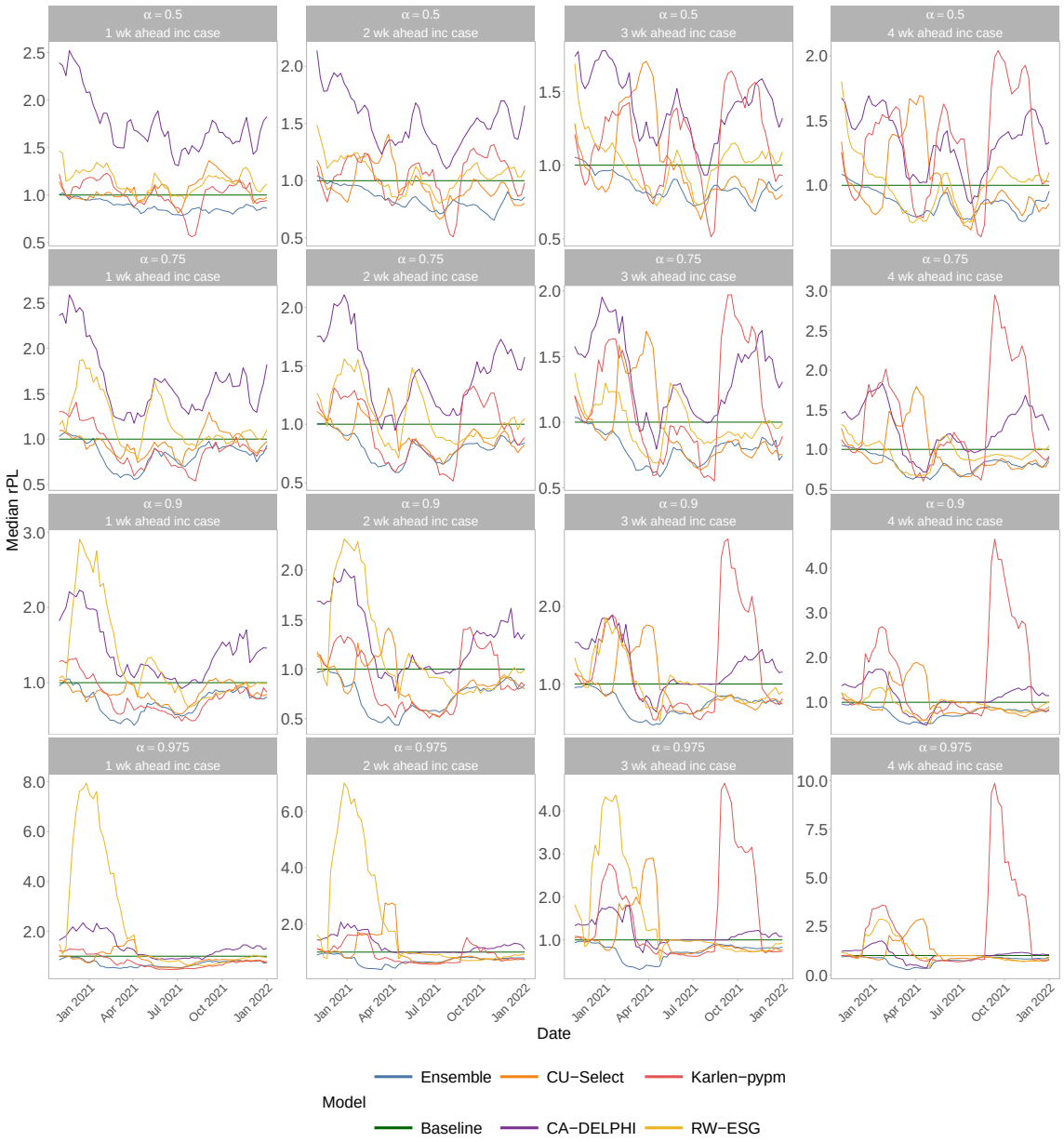


Figure SI G.16: **Median relative pinball loss (rPL) for each model's forecasts across individual locations from 12-week trailing windows:** Focusing on decision-makers with high C/L ratios (i.e., cost of action is expensive relative to harm), median rPL across all states (y-axis) is visualised for each model and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the trailing window.

rPL for the ensemble model's one-week-ahead forecasts at low and high C/L ratios

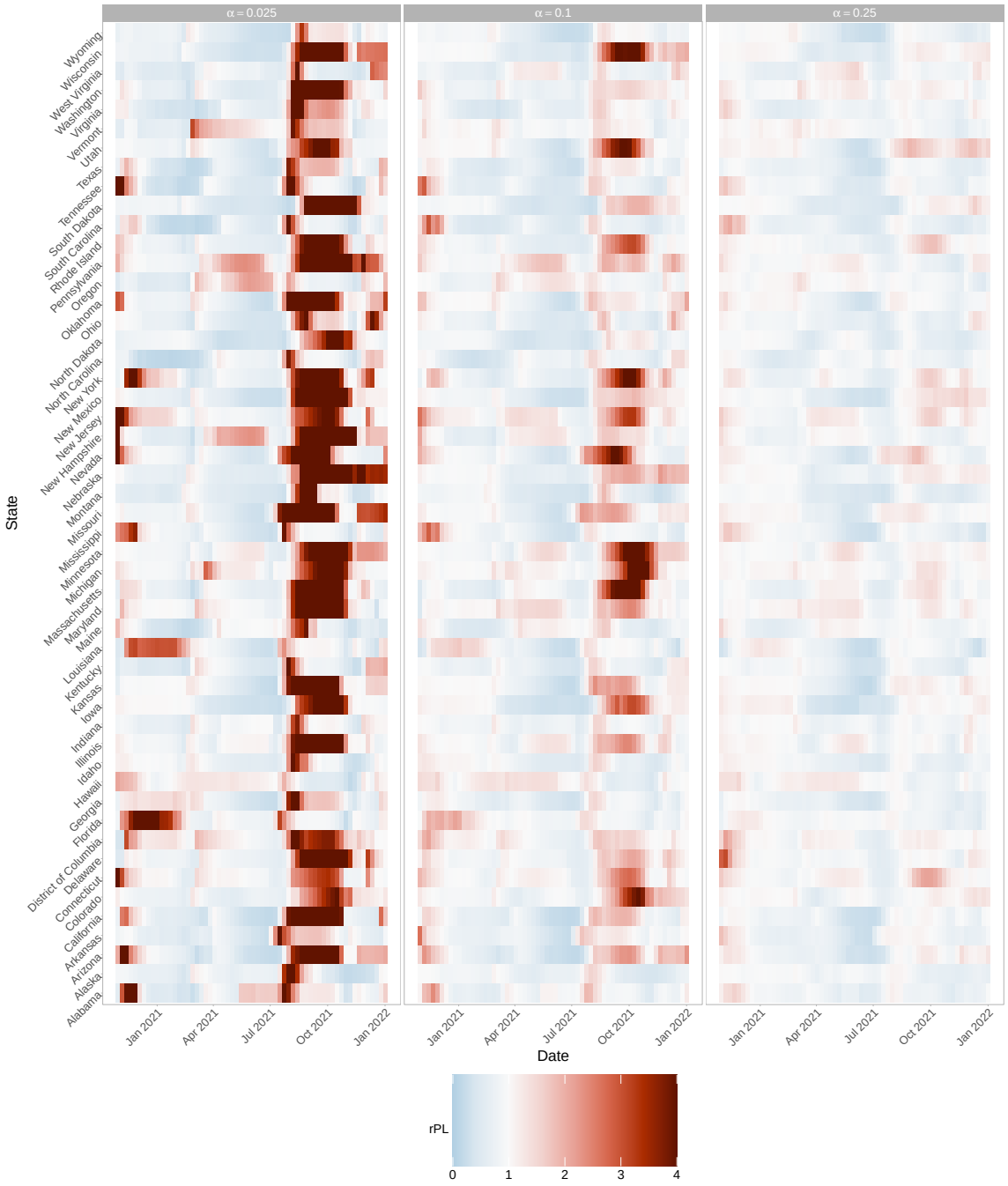


Figure SI G.17: **The relative pinball loss (rPL) for the ensemble's one-week-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on risk-averse decision-makers with low C/L ratios (i.e., cost of action is inexpensive relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

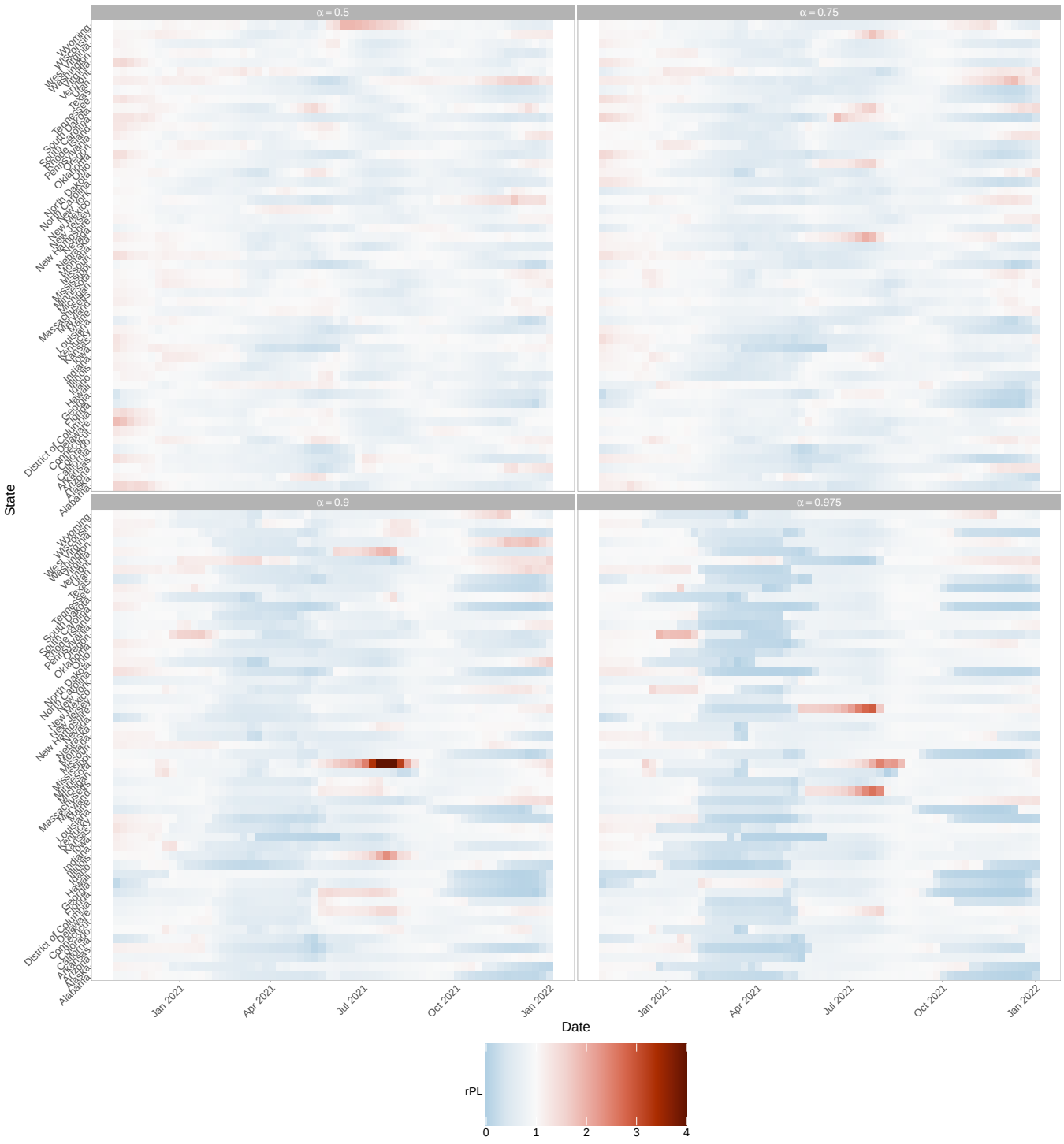


Figure SI G.18: **The relative pinball loss (rPL) for the ensemble's one-week-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on risk-tolerant decision-makers with high C/L ratios (i.e., cost of action is high relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

rPL for the ensemble model's two-weeks-ahead forecasts at low and high C/L ratios

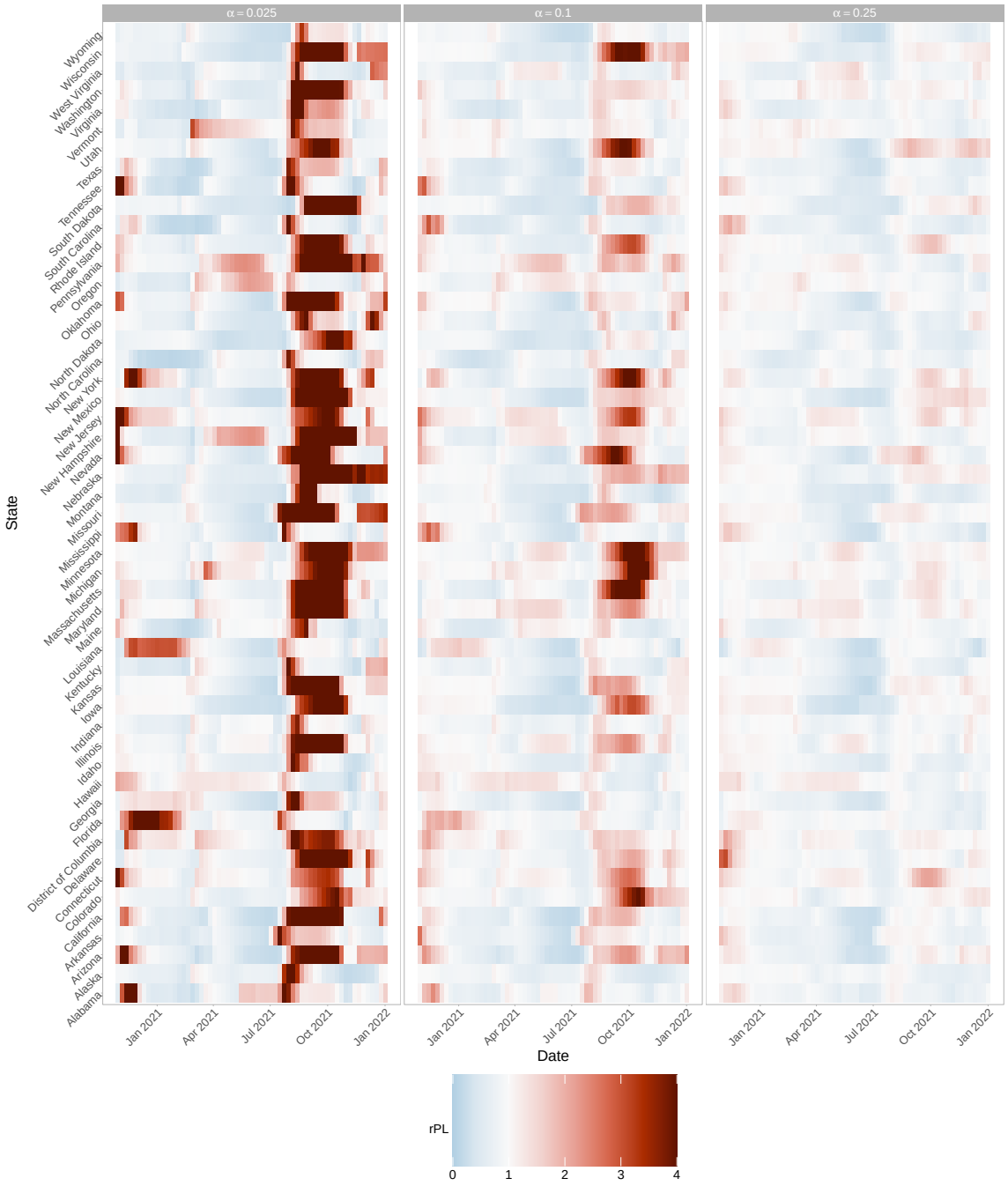


Figure SI G.19: **The relative pinball loss (rPL) for the ensemble's two-weeks-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on risk-averse decision-makers with low C/L ratios (i.e., cost of action is inexpensive relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

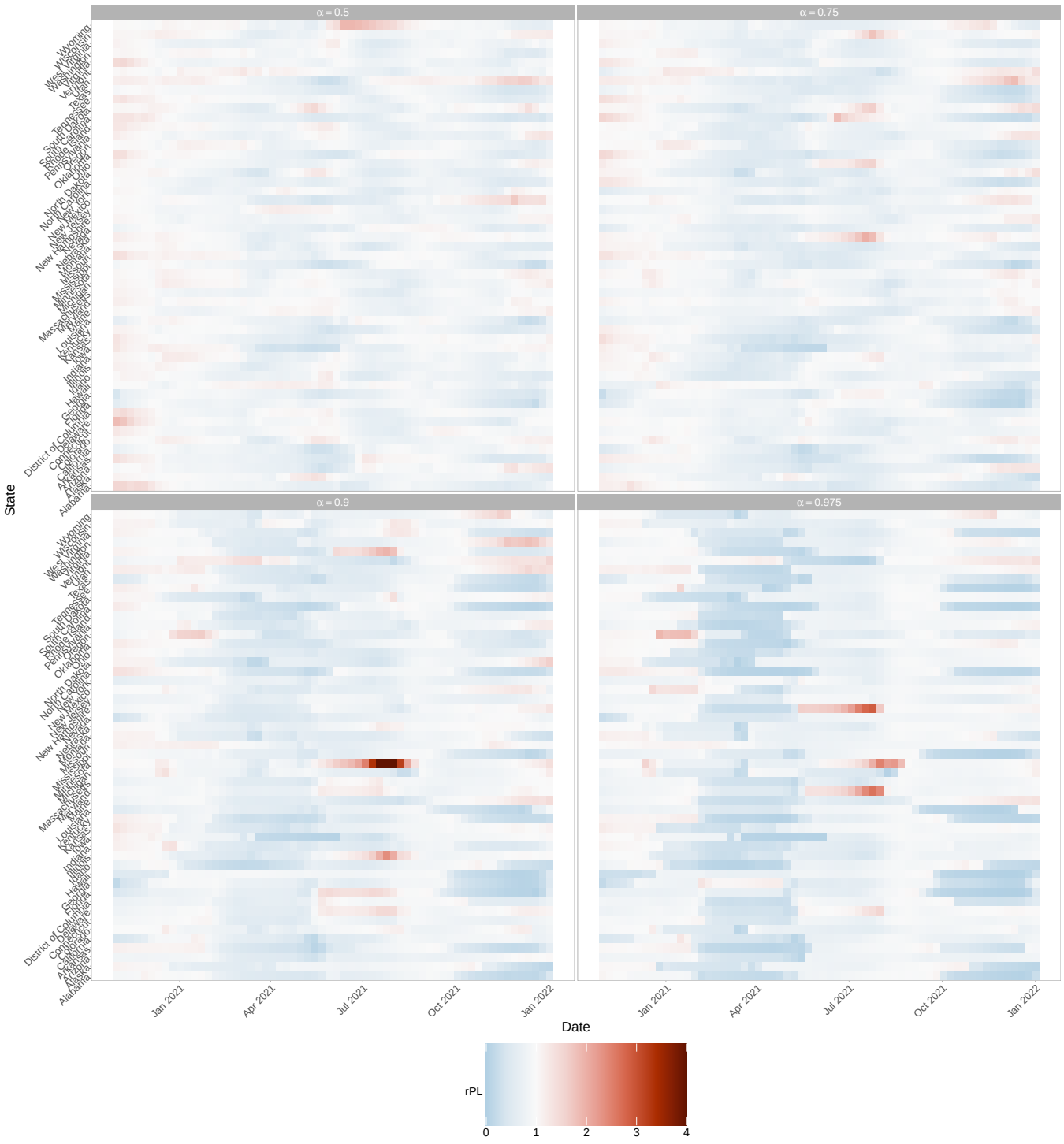


Figure SI G.20: **The relative pinball loss (rPL) for the ensemble's two-weeks-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on risk-tolerant decision-makers with high C/L ratios (i.e., cost of action is high relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

rPL for the ensemble model's three-weeks-ahead forecasts at low and high C/L ratios

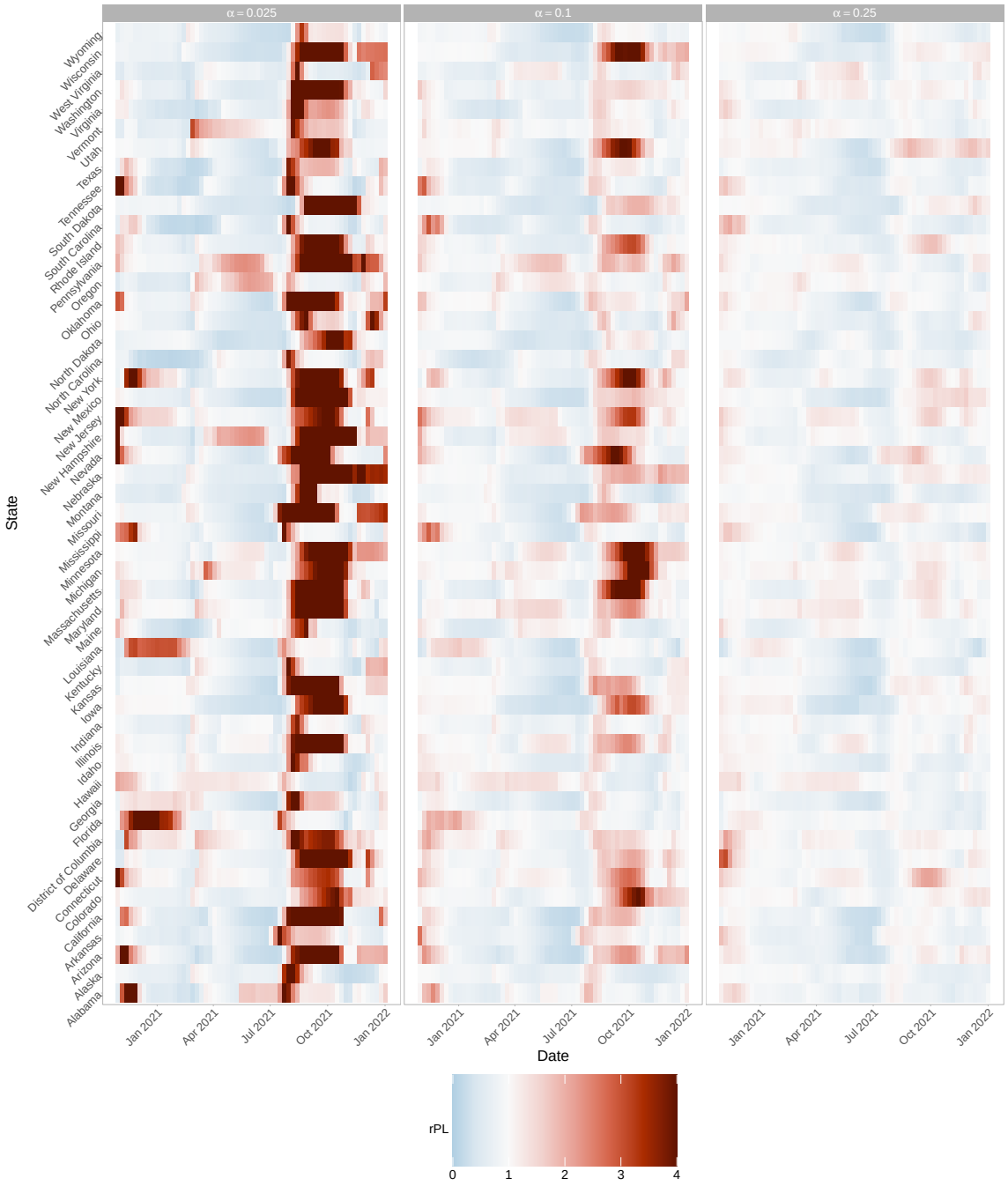


Figure SI G.21: **The relative pinball loss (rPL) for the ensemble's two-weeks-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on risk-averse decision-makers with low C/L ratios (i.e., cost of action is inexpensive relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

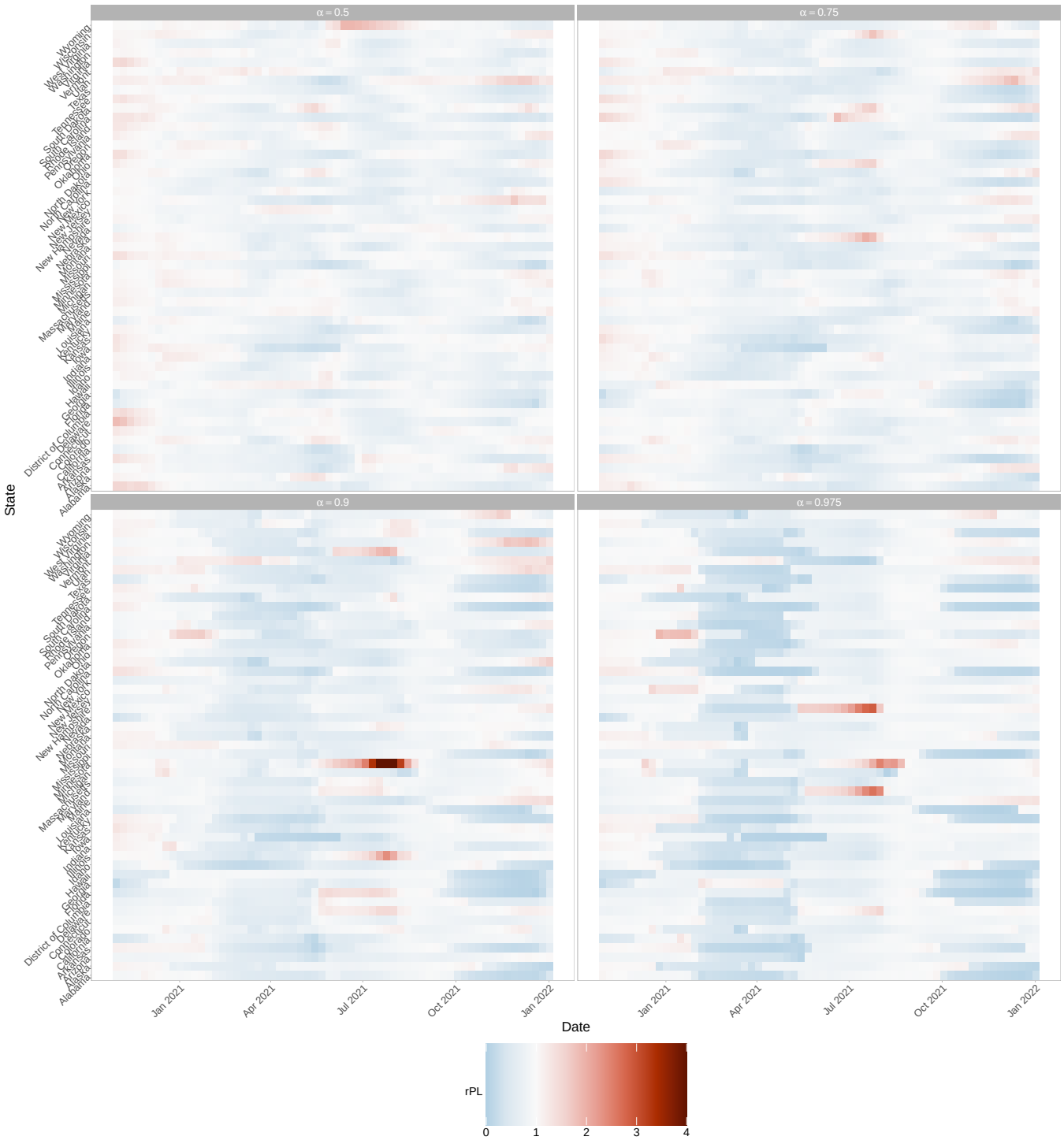


Figure SI G.22: **The relative pinball loss (rPL) for the ensemble's three-weeks-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on risk-tolerant decision-makers with high C/L ratios (i.e., cost of action is high relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

rPL for the ensemble model's four-weeks-ahead forecasts at low and high C/L ratios

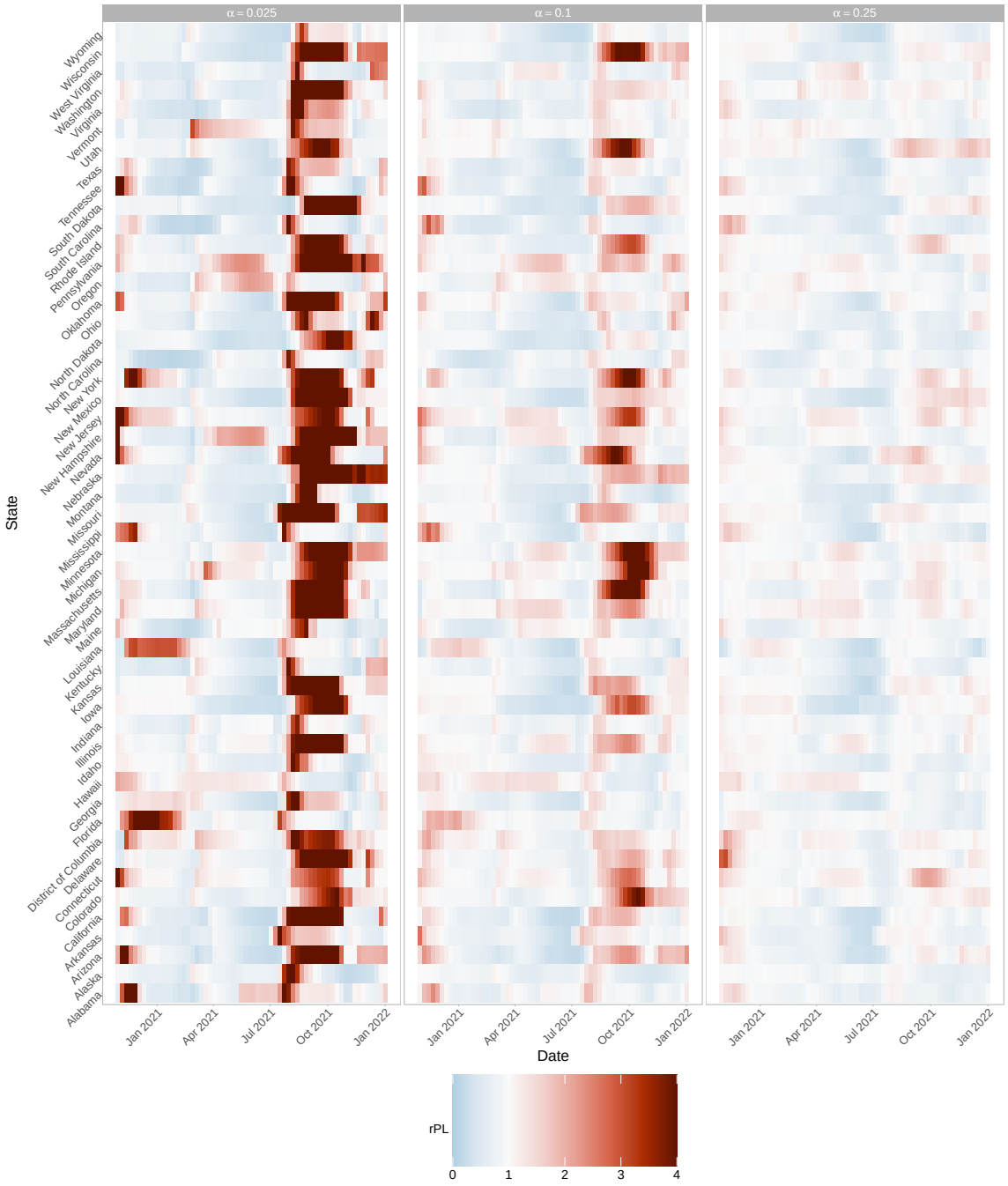


Figure SI G.23: **The relative pinball loss (rPL) for the ensemble's four-weeks-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on risk-averse decision-makers with low C/L ratios (i.e., cost of action is inexpensive relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

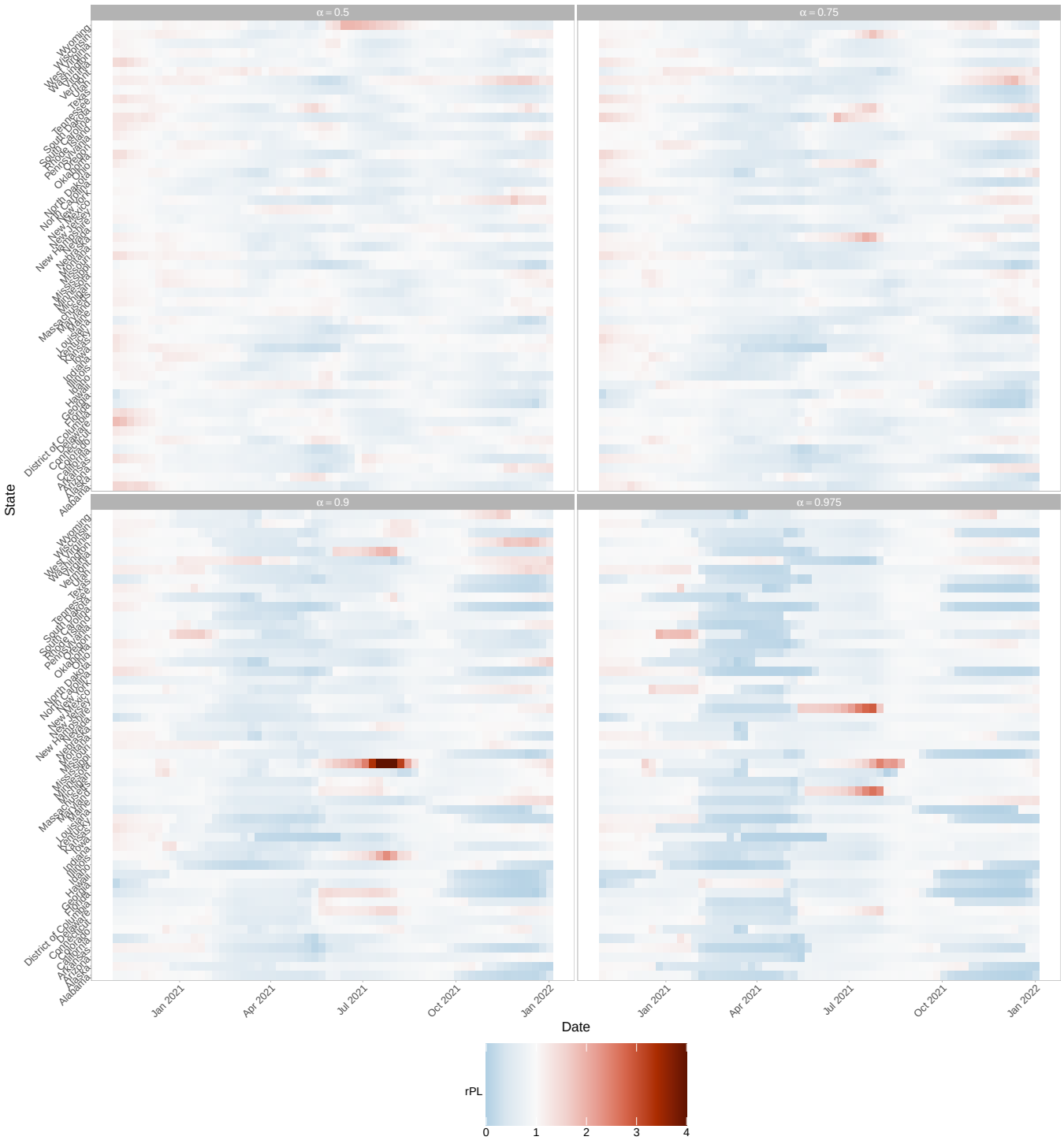


Figure SI G.24: **The relative pinball loss (rPL) for the ensemble's four-weeks-ahead forecasts for individual locations from 12-week trailing windows:** Focusing on decision-makers with high C/L ratios (i.e., cost of action is high relative to harm), rPL is visualised for each state (y-axis) and each trailing window where the x-axis denotes the maximum date of the 12-week trailing window. We calculated rPL relative to the baseline model, and rPL > 1 indicates greater forecast value for that decision-maker in the given location and trailing window.

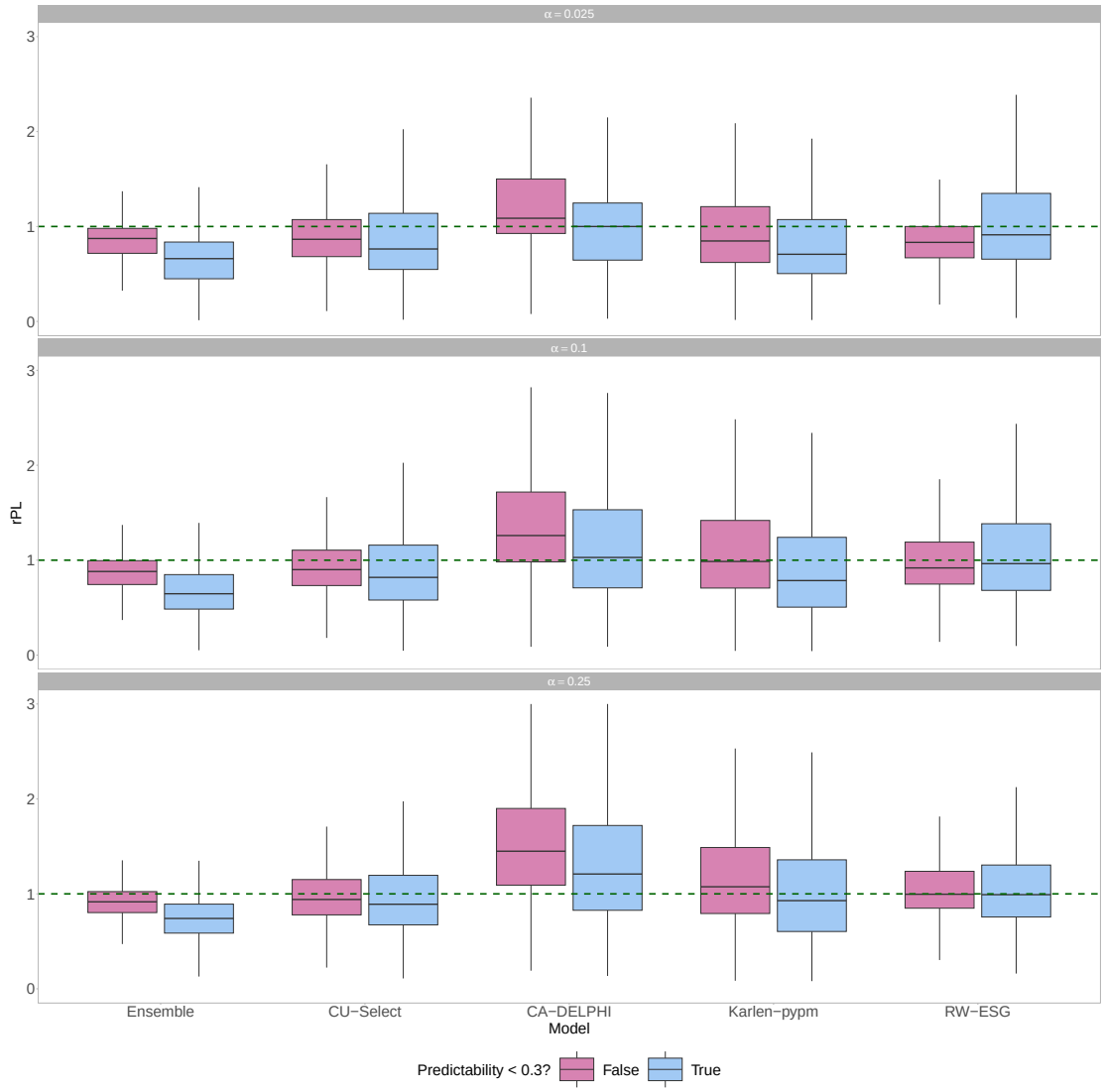


Figure SI G.25: **Relative pinball loss summary by low/high predictability regimes from trailing windows:** For visualisation, we use a binary classification here of predictability < 0.3 and focus on risk-intolerant decision-makers. Values of relative pinball loss (rPL) below 1 indicate greater forecast value at that fixed C/L ratio across all observations compared to the baseline model.

G.5.4 User-specific Murphy diagrams for higher C/L ratios

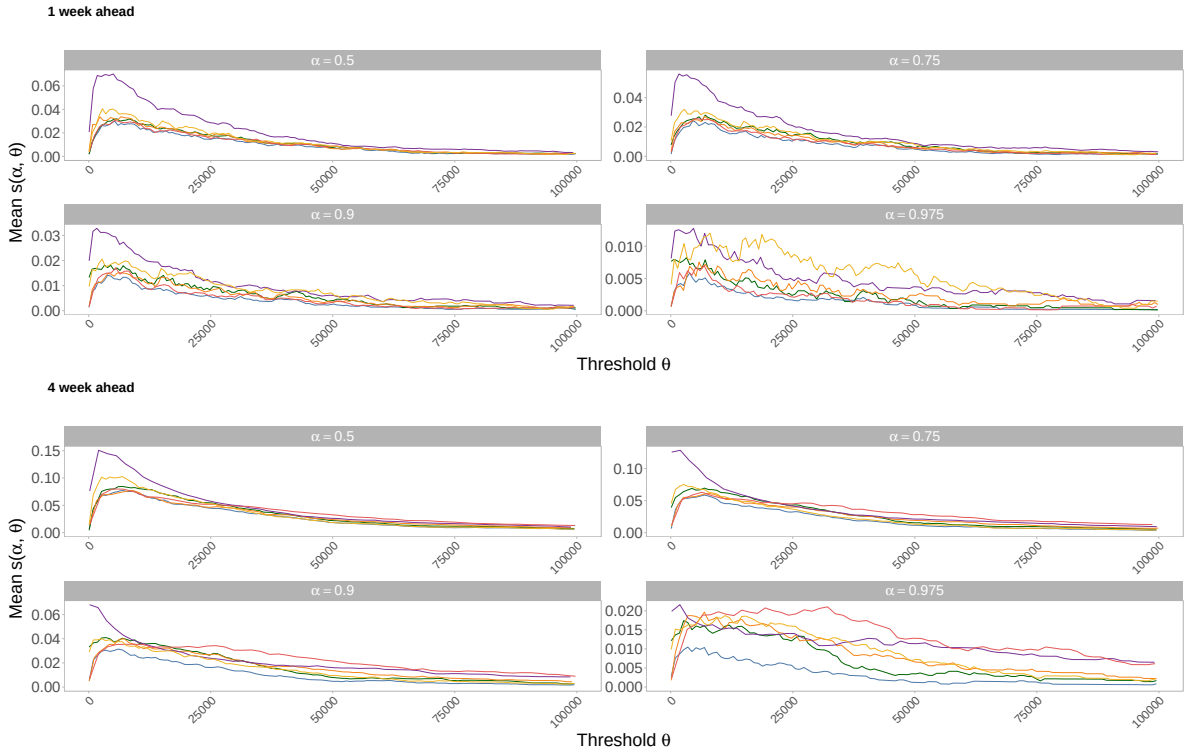


Figure SI G.26: **User-specific Murphy diagrams for higher C/L ratios:** We visualise Murphy diagrams for risk-tolerant decision-makers for whom the defined action is expensive relative to the loss/harm (i.e., high C/L ratios). Each Murphy curve captures mean elementary score $s(\alpha, \theta)$ for a fixed C/L ratio (α) and varying event thresholds (θ). Lower values on Murphy curves indicate superior performance of a model's forecasts for the given user-event combination. Figure 8.3 C.1 is the analogous visualisation for risk-intolerant decision-makers for whom the defined action is inexpensive relative to the loss/harm (i.e., low C/L ratios).

G.5.5 Predictability versus forecast metrics

Predictability versus R^2

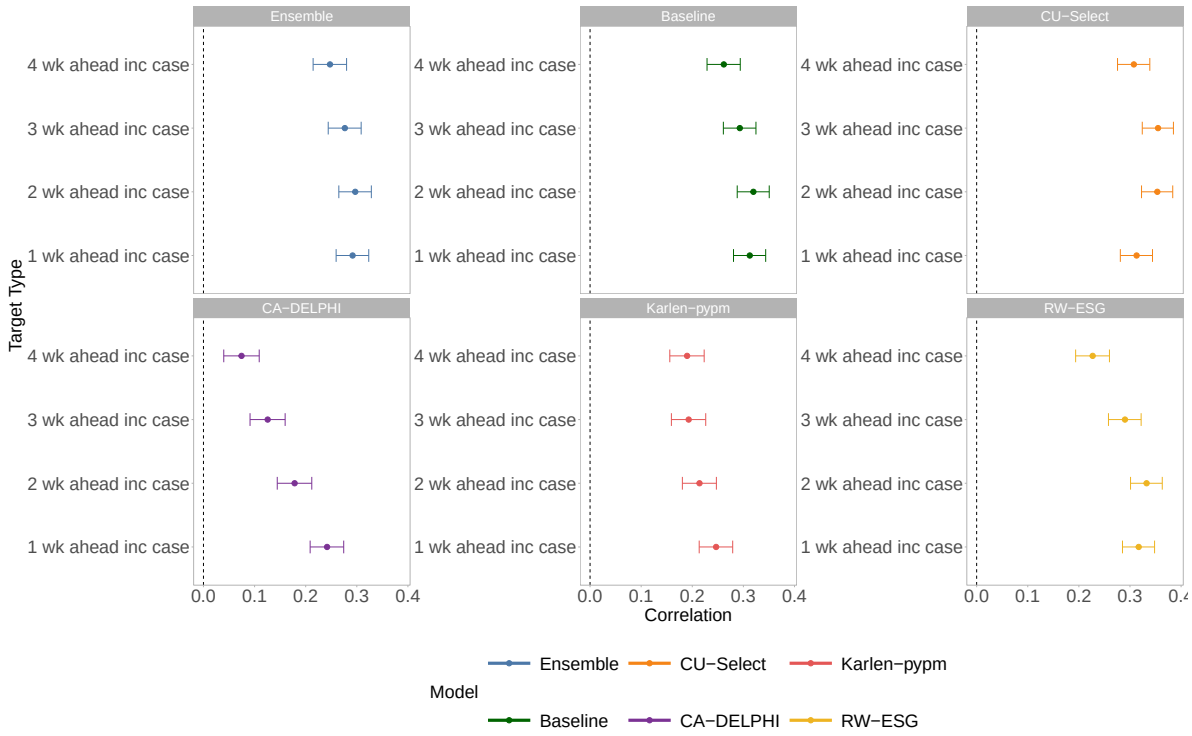


Figure SI G.27: **Correlations between predictability and R^2:** We visualise the Pearson's correlations (Table SI G.3) between predictability and R^2 for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

Table SI G.3: **Correlations between predictability and R^2:** For each model and forecasted quantity, we report the Pearson's[586] correlation r between predictability and R^2, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	0.29	[0.26, 0.32]	P < 0.001
COVIDhub-baseline	1-week-ahead	0.31	[0.28, 0.34]	P < 0.001
CU-select	1-week-ahead	0.31	[0.28, 0.34]	P < 0.001
CovidAnalytics-DELPHI	1-week-ahead	0.24	[0.21, 0.27]	P < 0.001
Karlen-pypm	1-week-ahead	0.25	[0.21, 0.28]	P < 0.001
RobertWalraven-ESG	1-week-ahead	0.32	[0.29, 0.35]	P < 0.001
Ensemble	2-weeks-ahead	0.30	[0.26, 0.33]	P < 0.001
COVIDhub-baseline	2-weeks-ahead	0.32	[0.29, 0.35]	P < 0.001
CU-select	2-weeks-ahead	0.35	[0.32, 0.38]	P < 0.001
CovidAnalytics-DELPHI	2-weeks-ahead	0.18	[0.14, 0.21]	P < 0.001
Karlen-pypm	2-weeks-ahead	0.21	[0.18, 0.25]	P < 0.001
RobertWalraven-ESG	2-weeks-ahead	0.33	[0.30, 0.36]	P < 0.001
Ensemble	3-weeks-ahead	0.28	[0.24, 0.31]	P < 0.001
COVIDhub-baseline	3-weeks-ahead	0.29	[0.26, 0.32]	P < 0.001
CU-select	3-weeks-ahead	0.35	[0.32, 0.39]	P < 0.001
CovidAnalytics-DELPHI	3-weeks-ahead	0.13	[0.09, 0.16]	P < 0.001
Karlen-pypm	3-weeks-ahead	0.19	[0.16, 0.23]	P < 0.001
RobertWalraven-ESG	3-weeks-ahead	0.29	[0.26, 0.32]	P < 0.001
Ensemble	4-weeks-ahead	0.25	[0.21, 0.28]	P < 0.001
COVIDhub-baseline	4-weeks-ahead	0.26	[0.23, 0.29]	P < 0.001
CU-select	4-weeks-ahead	0.31	[0.28, 0.34]	P < 0.001
CovidAnalytics-DELPHI	4-weeks-ahead	0.07	[0.04, 0.11]	P < 0.001
Karlen-pypm	4-weeks-ahead	0.19	[0.16, 0.22]	P < 0.001
RobertWalraven-ESG	4-weeks-ahead	0.23	[0.19, 0.26]	P < 0.001

Predictability versus 95% Prediction Interval (PI) Coverage

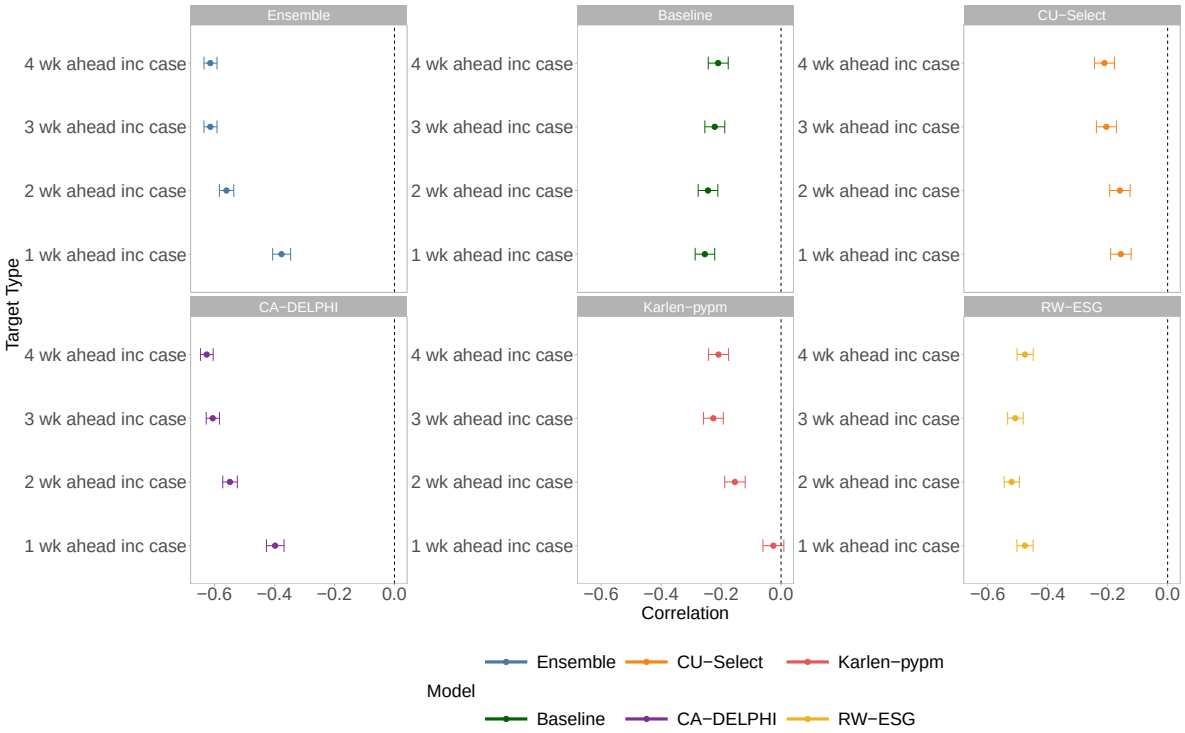


Figure SI G.28: **Correlations between predictability and central 95% prediction interval (PI) coverage:** We visualise the Pearson's correlations (Table SI G.4) between predictability and 95% PI coverage for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

Table SI G.4: **Correlations between predictability and central 95% Prediction Interval (PI) coverage:** For each model and forecasted quantity, we report the Pearson's[586] correlation r between predictability and 95% PI, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	-0.38	[-0.41, -0.35]	P < 0.001
COVIDhub-baseline	1-week-ahead	-0.25	[-0.29, -0.22]	P < 0.001
CU-select	1-week-ahead	-0.16	[-0.19, -0.12]	P < 0.001
CovidAnalytics-DELPHI	1-week-ahead	-0.40	[-0.43, -0.37]	P < 0.001
Karlen-pypm	1-week-ahead	-0.03	[-0.06, 0.01]	P = 1
RobertWalraven-ESG	1-week-ahead	-0.48	[-0.50, -0.45]	P < 0.001
Ensemble	2-weeks-ahead	-0.56	[-0.58, -0.54]	P < 0.001
COVIDhub-baseline	2-weeks-ahead	-0.24	[-0.28, -0.21]	P < 0.001
CU-select	2-weeks-ahead	-0.16	[-0.19, -0.13]	P < 0.001
CovidAnalytics-DELPHI	2-weeks-ahead	-0.55	[-0.57, -0.52]	P < 0.001
Karlen-pypm	2-weeks-ahead	-0.15	[-0.19, -0.12]	P < 0.001
RobertWalraven-ESG	2-weeks-ahead	-0.52	[-0.55, -0.49]	P < 0.001
Ensemble	3-weeks-ahead	-0.61	[-0.64, -0.59]	P < 0.001
COVIDhub-baseline	3-weeks-ahead	-0.22	[-0.25, -0.19]	P < 0.001
CU-select	3-weeks-ahead	-0.20	[-0.24, -0.17]	P < 0.001
CovidAnalytics-DELPHI	3-weeks-ahead	-0.61	[-0.63, -0.58]	P < 0.001
Karlen-pypm	3-weeks-ahead	-0.23	[-0.26, -0.19]	P < 0.001
RobertWalraven-ESG	3-weeks-ahead	-0.51	[-0.53, -0.48]	P < 0.001
Ensemble	4-weeks-ahead	-0.61	[-0.64, -0.59]	P < 0.001
COVIDhub-baseline	4-weeks-ahead	-0.21	[-0.24, -0.18]	P < 0.001
CU-select	4-weeks-ahead	-0.21	[-0.24, -0.18]	P < 0.001
CovidAnalytics-DELPHI	4-weeks-ahead	-0.63	[-0.65, -0.60]	P < 0.001
Karlen-pypm	4-weeks-ahead	-0.21	[-0.24, -0.18]	P < 0.001
RobertWalraven-ESG	4-weeks-ahead	-0.48	[-0.50, -0.45]	P < 0.001

Predictability versus relative Weighted Interval Score (rWIS)

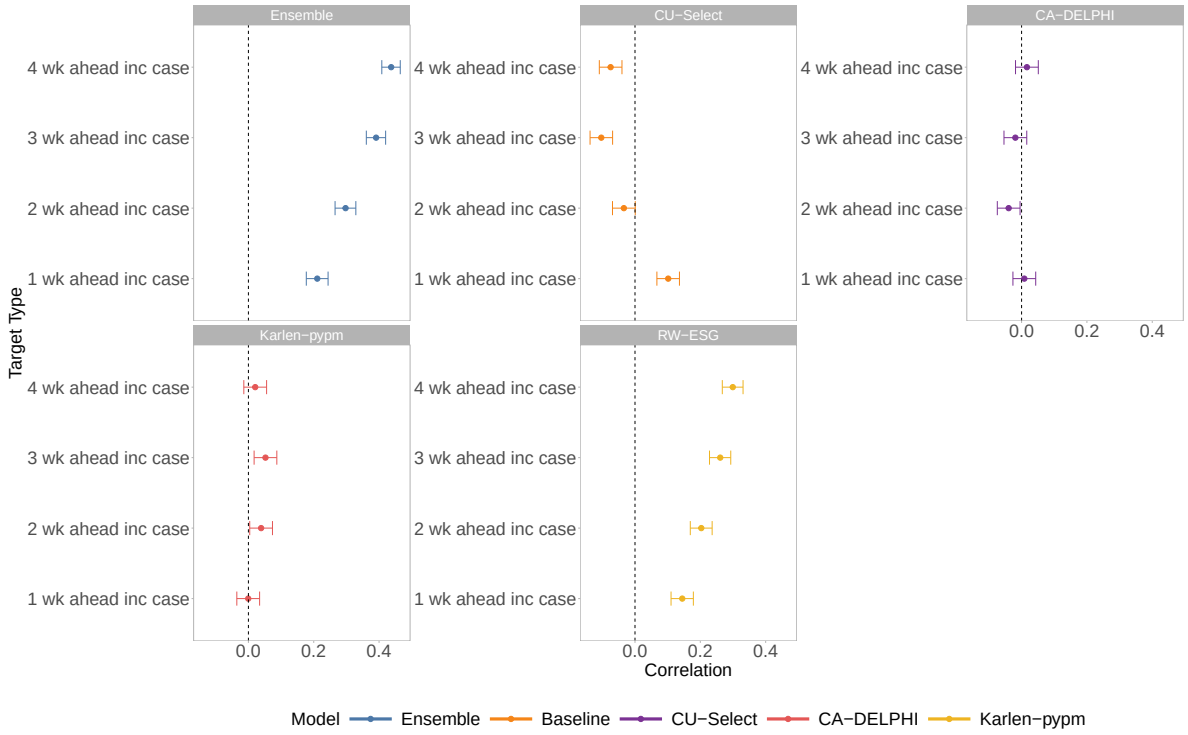


Figure SI G.29: **Correlations between predictability and relative Weighted Interval Score (rWIS)** We visualise the Pearson's correlations (Table SI G.5) between predictability and rWIS for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

Table SI G.5: **Correlations between predictability and relative Weighted Interval Score (rWIS)** For each model and forecasted quantity, we report the Pearson's [586] correlation r between predictability and rWIS, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$. The baseline values are intentionally marked as NA because it is the comparison model for rWIS.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	0.21	[0.18, 0.24]	P < 0.001
COVIDhub-baseline	1-week-ahead	NA	NA	NA
CU-select	1-week-ahead	0.10	[0.07, 0.14]	P < 0.001
CovidAnalytics-DELPHI	1-week-ahead	0.01	[-0.03, 0.04]	P = 1
Karlen-pypm	1-week-ahead	-0.00	[-0.04, 0.03]	P = 1
RobertWalraven-ESG	1-week-ahead	0.14	[0.11, 0.18]	P < 0.001
Ensemble	2-weeks-ahead	0.30	[0.27, 0.33]	P < 0.001
COVIDhub-baseline	2-weeks-ahead	NA	NA	NA
CU-select	2-weeks-ahead	-0.03	[-0.07, 0.00]	P = 1
CovidAnalytics-DELPHI	2-weeks-ahead	-0.04	[-0.07, -0.00]	P = 0.525
Karlen-pypm	2-weeks-ahead	0.04	[0.00, 0.07]	P = 0.574
RobertWalraven-ESG	2-weeks-ahead	0.20	[0.17, 0.24]	P < 0.001
Ensemble	3-weeks-ahead	0.39	[0.36, 0.42]	P < 0.001
COVIDhub-baseline	3-weeks-ahead	NA	NA	NA
CU-select	3-weeks-ahead	-0.10	[-0.14, -0.07]	P < 0.001
CovidAnalytics-DELPHI	3-weeks-ahead	-0.02	[-0.05, 0.02]	P = 1
Karlen-pypm	3-weeks-ahead	0.05	[0.02, 0.09]	P = 0.0658
RobertWalraven-ESG	3-weeks-ahead	0.26	[0.23, 0.29]	P < 0.001
Ensemble	4-weeks-ahead	0.44	[0.41, 0.46]	P < 0.001
COVIDhub-baseline	4-weeks-ahead	NA	NA	NA
CU-select	4-weeks-ahead	-0.07	[-0.11, -0.04]	P < 0.001
CovidAnalytics-DELPHI	4-weeks-ahead	0.02	[-0.02, 0.05]	P = 1
Karlen-pypm	4-weeks-ahead	0.02	[-0.01, 0.06]	P = 1
RobertWalraven-ESG	4-weeks-ahead	0.30	[0.27, 0.33]	P < 0.001

G.5.6 Predictability and $R(t)$

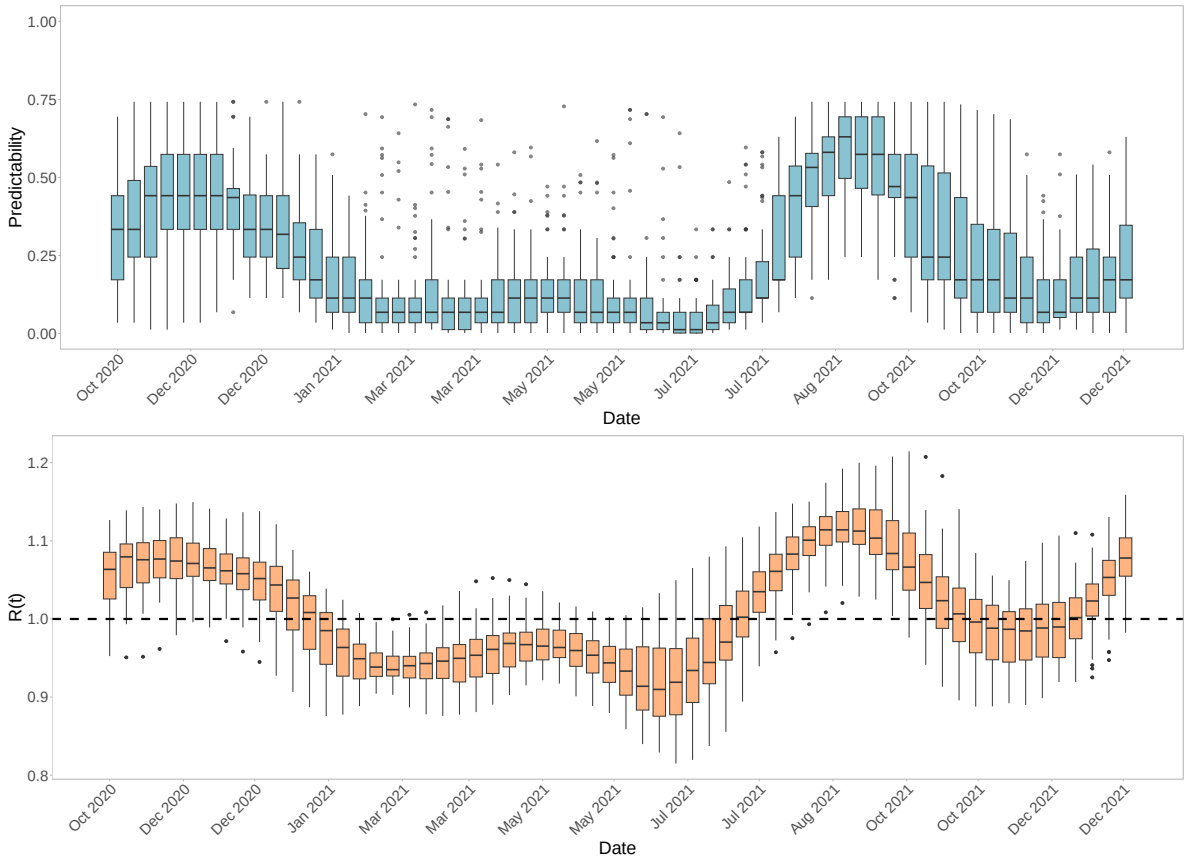


Figure SI G.30: **Predictability and $R(t)$:** We visualise the boxplots of window-specific predictability (top row) and mean time-varying reproduction number $R(t)$ (bottom row), which summarise the values (y-axis) obtained in each state in each 12-week trailing window, where the maximum date of each window is shown (x-axis). The box and whiskers cover the interquartile range, and points represent outliers.

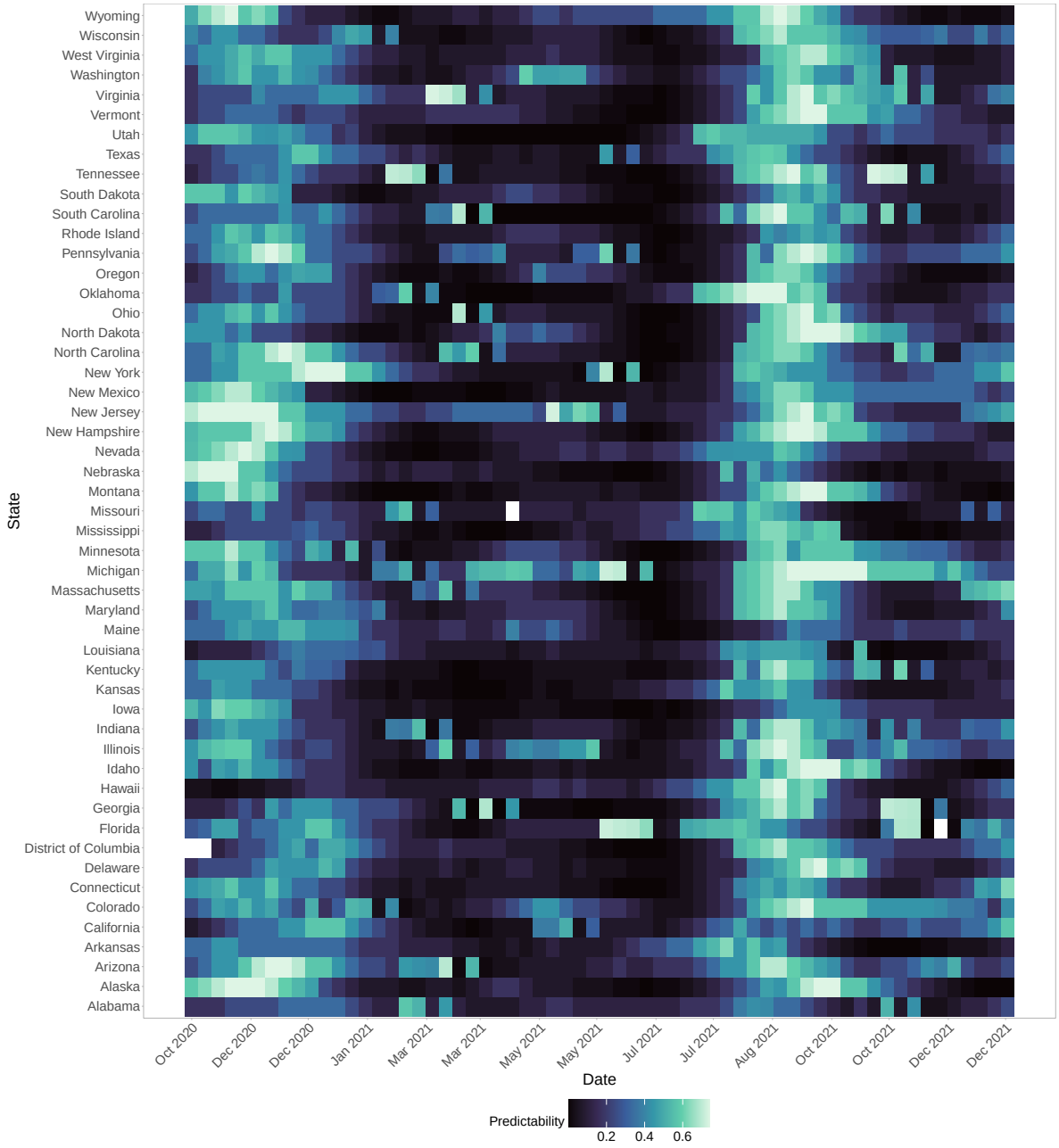


Figure SI G.31: **Predictability over time and across space:** We visualise the window-specific predictability of each state (y-axis) in each 12-week trailing window, where the maximum date of each window is shown (x-axis).

G.6 COVID-19 wave further analyses

G.6.1 Pre-wave results

Before the wave, predictability had been high during a previous wave and low at recent low case levels (Figure 8.4 B). High predictability had been moderately-to-strongly correlated with higher $R(t)$ ($r = 0.57$, 95% CI: 0.29, 0.76). Based on overall predictive performance (rWIS, Figure 8.4 B), the ensemble and Karlen-pypm forecasting models were likely the most suitable forecasting models to use going forward. Forecast value for specific events was generally higher for these models when measured using the BS, with superior discrimination and calibration properties (Figure SI G.32 A), yet the "best" model depended on different decision-maker risk preferences (via C/L ratios), event thresholds, and forecast horizon (Figures SI G.35 A, event-specific Murphy diagrams of Figure SI G.33 A, or user-specific Murphy diagrams of Figure SI G.34 A). The ensemble model generally provided the most value before the wave for one-week-ahead forecasts across all event thresholds for different decision-maker preferences (Figure SI G.35 A), yet for four-weeks-ahead forecasts, the RW-ESG and Karlen-pypm model provided more clear value for risk-tolerant (high C/L ratio) and risk-intolerant (low C/L ratio) respectively (Figure SI G.35 A). Compared to the baseline model, the ensemble provided the most consistent clear relative value for both risk-intolerant (Figure SI G.36) and risk-tolerant (Figure SI G.37) over time.

We found similar results for the pre-wave period, compared to the entire studied period (Section 8.6.2), with many models having higher R^2 (Figures SI G.38, SI G.40, Table SI G.3) and lower 95% PI coverage (Figures SI G.38, SI G.42, Table SI G.8) during times of greater predictability, which is what marks the onset of this wave, Figure 8.4 A). During such high predictability regimes, the ensemble offered value over the baseline model for risk-tolerant and risk-intolerant decision-makers (pinball loss, Figures SI G.38, SI G.36–SI G.37), yet offered comparably less value (over the baseline) during times of lower predictability (rWIS and relative pinball loss, Figure SI G.38). Before the wave, only the Karlen-pypm model had regularly outperformed the baseline in terms of overall predictive performance (rWIS) across both high and low predictability regimes, yet its relative value varied between risk-intolerant and risk-tolerant decision-makers.

G.6.2 Brier score decomposition before and during wave

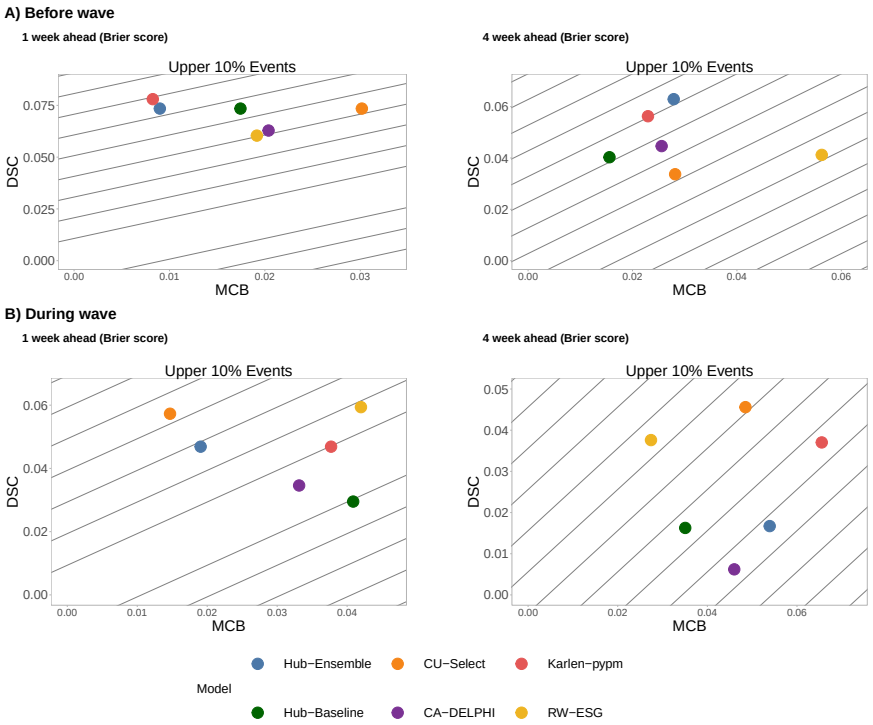


Figure SI G.32: **Brier score decomposition for the upper 10% of cases before and during wave:** For the state of Arizona, we visualise the discrimination ability (DSC) component (y-axis) and miscalibration (MCB) component (x-axis) of each model's Brier score for the upper 10% of cases before ((**A**)) and during ((**B**)) the COVID-19 wave which started on 26 June 2021 (when the median $R(t)$ estimate was greater than 1). Higher DSC values and lower MCB values indicate greater forecast value for a the event (i.e., event probability/frequency $\pi(\theta) = 0.1$) across the set of observations.

G.6.3 Event-specific Murphy diagrams before and during wave

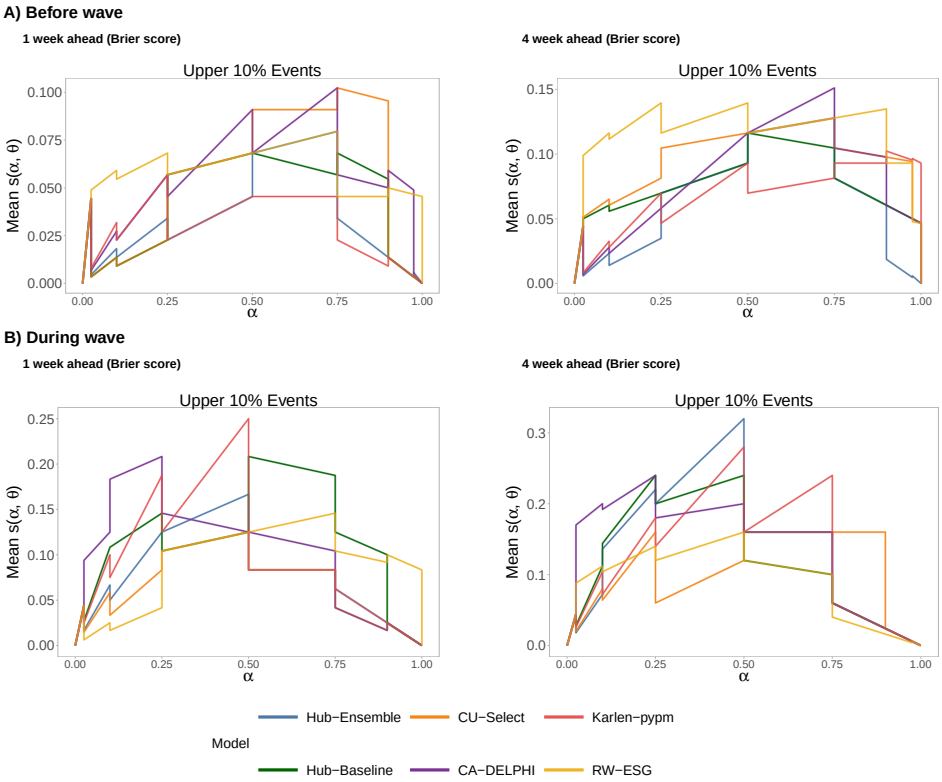


Figure SI G.33: **Event-specific Murphy diagrams for the upper 10% of cases before and during wave:** For the state of Arizona, we visualise the Murphy diagrams relevant to specific decision-makers before the studied wave. Each Murphy curve captures mean elementary score $s(\alpha, \theta)$ for a fixed event threshold (θ) and varying C/L ratios (α). Lower values on Murphy curves indicate superior performance of a model's forecasts for the given user-event combination.

G.6.4 User-specific Murphy diagrams before and during wave

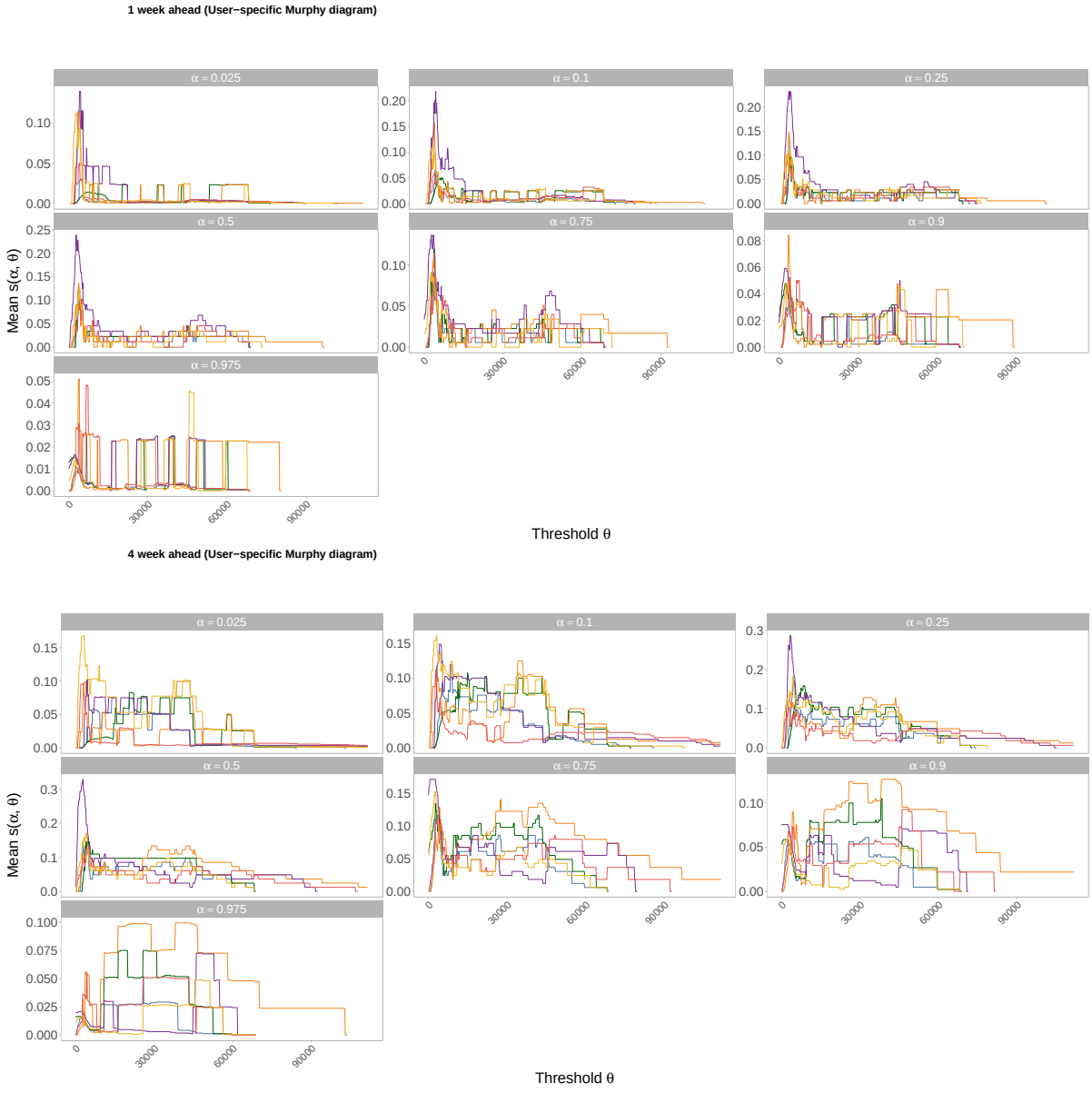


Figure SI G.34: **User-specific Murphy diagrams before wave:** For the state of Arizona, we visualise the Murphy diagrams relevant to specific decision-makers before the studied COVID-19 wave which started on 26 June 2021 (when the median $R(t)$ estimate was greater than 1). Each Murphy curve captures mean elementary score $s(\alpha, \theta)$ for a fixed C/L ratio (α) and varying event thresholds (θ). Lower values on Murphy curves indicate superior performance of a model's forecasts for the given user-event combination.

G.6.5 Pinball loss decomposition before and during wave

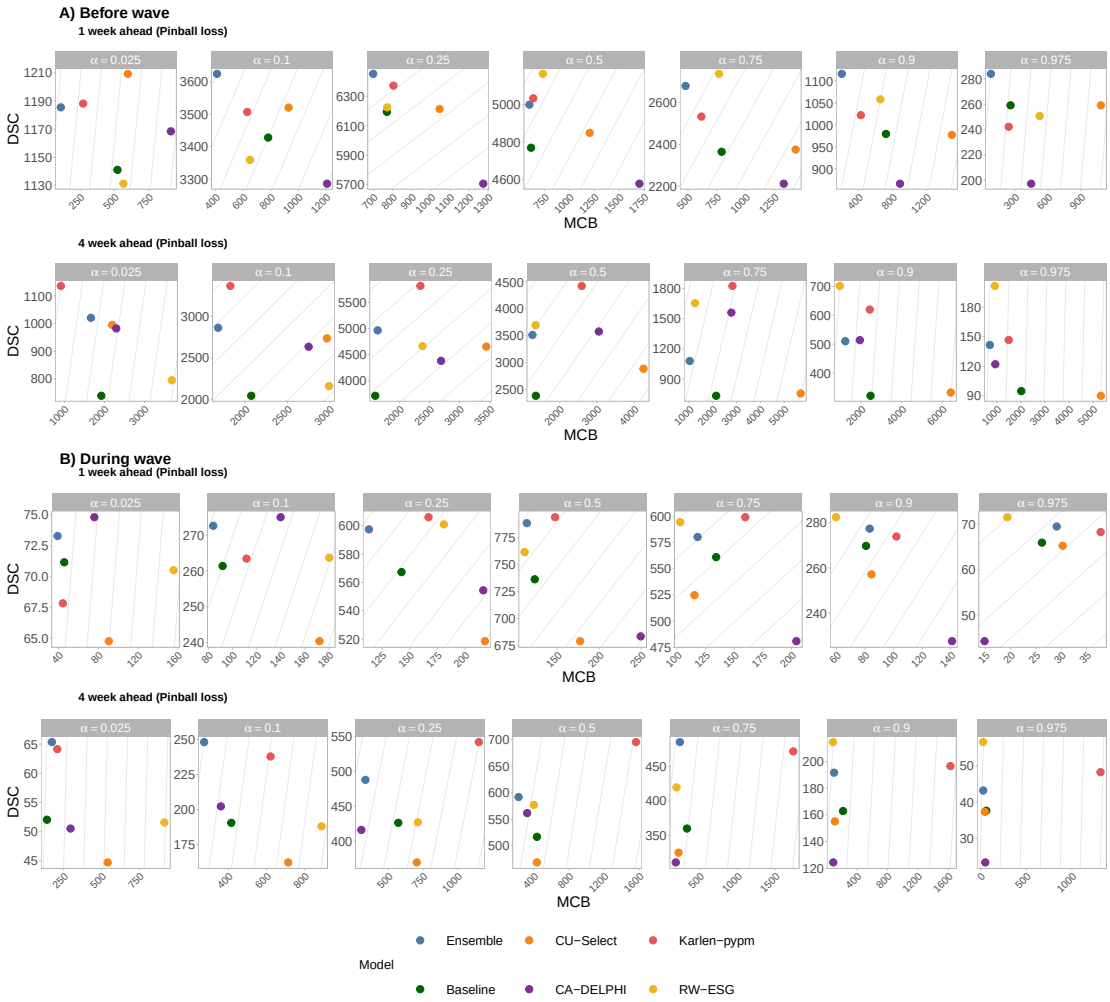


Figure SI G.35: **Pinball loss decomposition before and during wave:** For the state of Arizona, we visualise the discrimination ability (DSC) component (y-axis) and miscalibration (MCB) component (x-axis) of each model's pinball loss before ((**A**)) and during ((**B**)) the COVID-19 wave which started on 26 June 2021 (when the median $R(t)$ estimate was greater than 1). Higher DSC values and lower MCB values indicate greater forecast value for a specific decision-maker's C/L ratio α across the set of observations.

G.6.6 Relative pinball loss before and during wave

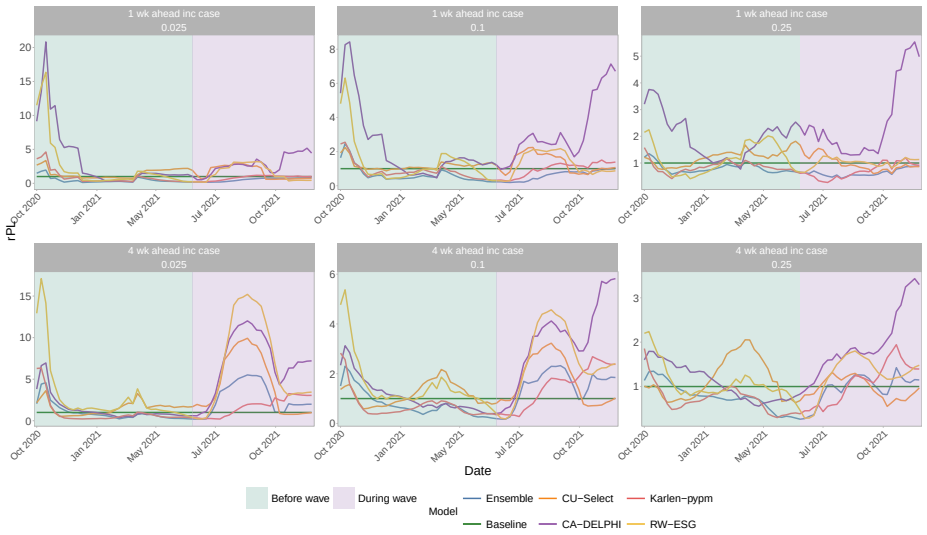


Figure SI G.36: **Relative pinball loss over time for risk-intolerant decision-makers before and during wave:** For the state of Arizona, we visualise each model's relative pinball loss (rPL) (y-axis) before and during the COVID-19 wave which started on 26 June 2021 (when the median $R(t)$ estimate was greater than 1). rPL was calculated using the COVID-19 Forecast Hub baseline model in each 12-week trailing window, where the maximum date of each window is shown (x-axis). Values of rPL < 1 indicate greater relative forecast value of a given model, compared to the baseline model, for observations across that 12-week window and specific decision-maker's preferences (C/L ratio, α). We focus on risk-tolerant decision-makers in Figure SI G.37

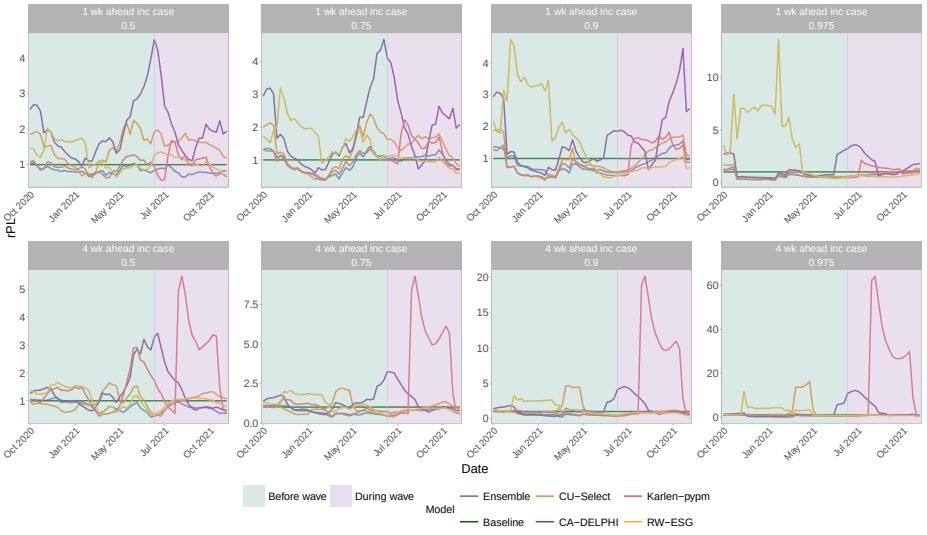


Figure SI G.37: **Relative pinball loss over time for risk-tolerant decision-makers before and during wave:** For the state of Arizona, we visualise each model's relative pinball loss (rPL) (y-axis) before and during the COVID-19 wave which started on 26 June 2021 (when the median $R(t)$ estimate was greater than 1). rPL was calculated using the COVID-19 Forecast Hub baseline model in each 12-week trailing window, where the maximum date of each window is shown (x-axis). Values of rPL < 1 indicate greater relative forecast value of a given model, compared to the baseline model, for observations across that 12-week window and specific decision-maker's preferences (C/L ratio, α). We focus on risk-intolerant (lower C/L ratio) decision-makers in Figure SI G.36

G.6.7 Predictability impacts before and during wave

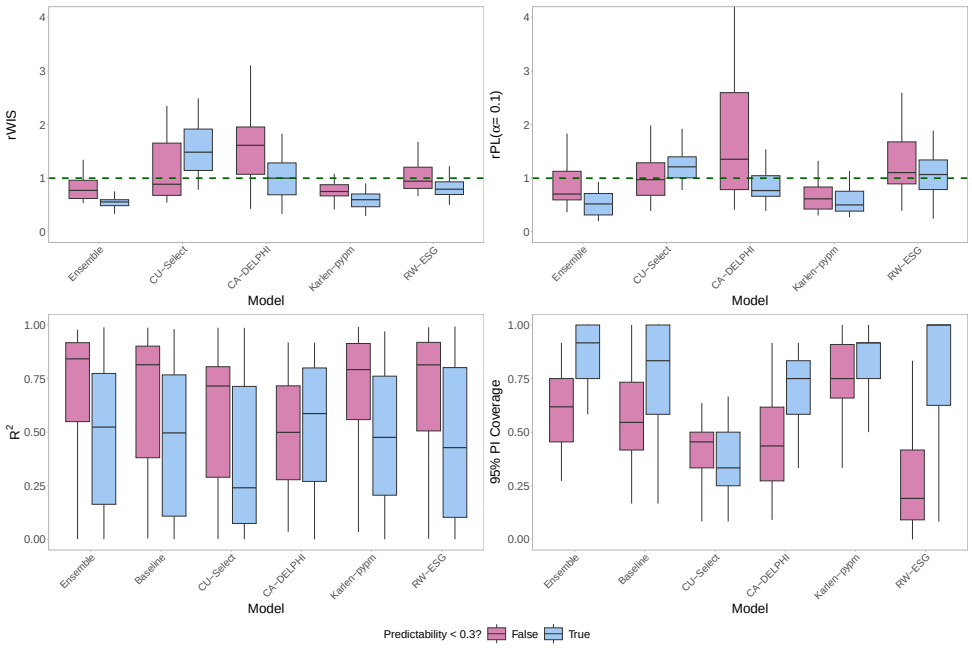


Figure SI G.38: **Pre-wave predictability and forecast performance:** For the state of Arizona, we visualise the relationship between forecast performance metrics and predictability before the onset of the studied wave (starting July 2021).

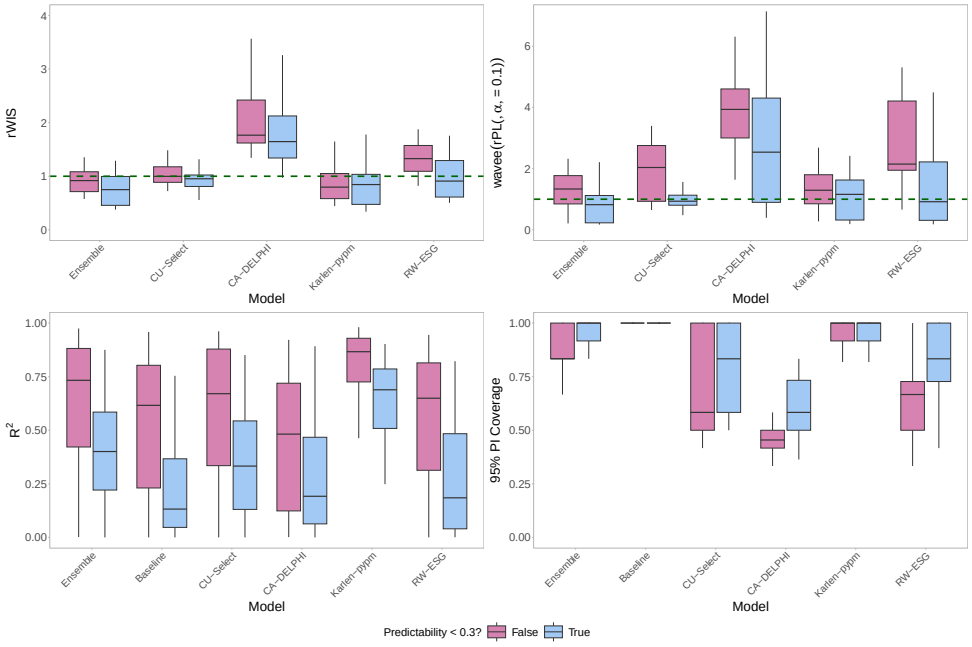


Figure SI G.39: **During-wave predictability and forecast performance:** For the state of Arizona, we visualise the relationship between forecast performance metrics and predictability during the studied wave (starting July 2021).

G.6.8 Predictability versus forecast metrics

Predictability versus R^2

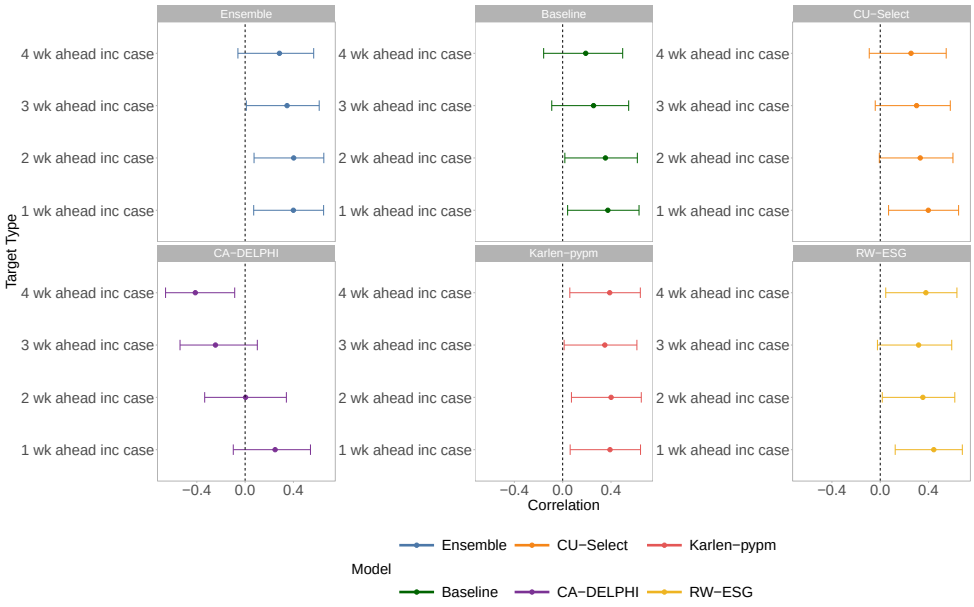


Figure SI G.40: **Pre-wave correlations between predictability and R^2:** In Arizona before the onset of the studied wave (starting July 2021), we visualise the Pearson's correlations (Table SI G.3) between predictability and R^2 for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

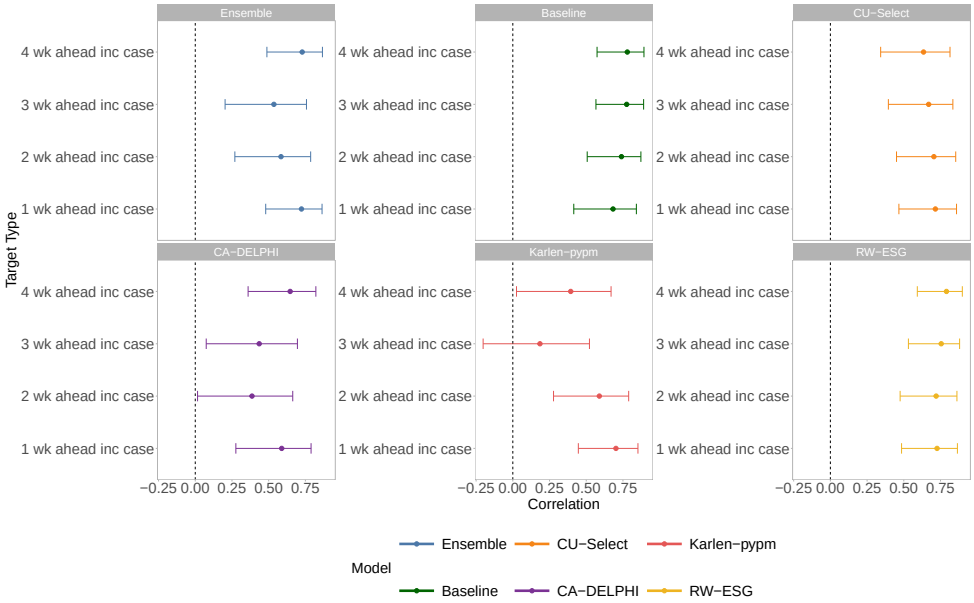


Figure SI G.41: **During-wave correlations between predictability and R^2:** In Arizona during the studied wave (starting July 2021), we visualise the Pearson's correlations (Table SI G.3) between predictability and R^2 for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

Table SI G.6: **Pre-wave correlations between predictability and R^2:** In Arizona before the onset of the studied wave (starting July 2021), for each model and forecasted quantity, we report the Pearson's[586] correlation r between predictability and R^2, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	0.40	[0.07, 0.65]	P = 0.461
Ensemble	2-weeks-ahead	0.40	[0.07, 0.65]	P = 0.442
Ensemble	3-weeks-ahead	0.35	[0.01, 0.61]	P = 1
Ensemble	4-weeks-ahead	0.28	[-0.06, 0.57]	P = 1
COVIDhub-baseline	1-week-ahead	0.37	[0.04, 0.63]	P = 0.703
COVIDhub-baseline	2-weeks-ahead	0.35	[0.02, 0.62]	P = 0.959
COVIDhub-baseline	3-weeks-ahead	0.26	[-0.09, 0.55]	P = 1
COVIDhub-baseline	4-weeks-ahead	0.19	[-0.16, 0.50]	P = 1
CU-select	1-week-ahead	0.40	[0.07, 0.65]	P = 0.473
CU-select	2-weeks-ahead	0.33	[-0.01, 0.60]	P = 1
CU-select	3-weeks-ahead	0.30	[-0.04, 0.58]	P = 1
CU-select	4-weeks-ahead	0.26	[-0.09, 0.55]	P = 1
CovidAnalytics-DELPHI	1-week-ahead	0.25	[-0.10, 0.54]	P = 1
CovidAnalytics-DELPHI	2-weeks-ahead	0.00	[-0.33, 0.34]	P = 1
CovidAnalytics-DELPHI	3-weeks-ahead	-0.25	[-0.54, 0.10]	P = 1
CovidAnalytics-DELPHI	4-weeks-ahead	-0.41	[-0.66, -0.09]	P = 0.369
Karlen-pypm	1-week-ahead	0.39	[0.06, 0.64]	P = 0.525
Karlen-pypm	2-weeks-ahead	0.40	[0.07, 0.65]	P = 0.451
Karlen-pypm	3-weeks-ahead	0.35	[0.01, 0.61]	P = 1
Karlen-pypm	4-weeks-ahead	0.39	[0.06, 0.64]	P = 0.546
RobertWalraven-ESG	1-week-ahead	0.44	[0.12, 0.68]	P = 0.207
RobertWalraven-ESG	2-weeks-ahead	0.35	[0.02, 0.62]	P = 0.974
RobertWalraven-ESG	3-weeks-ahead	0.32	[-0.02, 0.59]	P = 1
RobertWalraven-ESG	4-weeks-ahead	0.38	[0.05, 0.63]	P = 0.66

Table SI G.7: **During-wave correlations between predictability and R^2:** In Arizona before the onset of the studied wave (starting July 2021), for each model and forecasted quantity, we report the Pearson's[586] correlation r between predictability and R^2, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	0.72	[0.48, 0.86]	P < 0.001
Ensemble	2-weeks-ahead	0.58	[0.27, 0.79]	P = 0.0265
Ensemble	3-weeks-ahead	0.54	[0.20, 0.76]	P = 0.0794
Ensemble	4-weeks-ahead	0.73	[0.49, 0.87]	P < 0.001
COVIDhub-baseline	1-week-ahead	0.68	[0.42, 0.84]	P = 0.00151
COVIDhub-baseline	2-weeks-ahead	0.74	[0.51, 0.87]	P < 0.001
COVIDhub-baseline	3-weeks-ahead	0.78	[0.57, 0.89]	P < 0.001
COVIDhub-baseline	4-weeks-ahead	0.78	[0.57, 0.89]	P < 0.001
CU-select	1-week-ahead	0.72	[0.47, 0.86]	P < 0.001
CU-select	2-weeks-ahead	0.71	[0.45, 0.85]	P < 0.001
CU-select	3-weeks-ahead	0.67	[0.39, 0.83]	P = 0.00235
CU-select	4-weeks-ahead	0.63	[0.34, 0.81]	P = 0.00691
CovidAnalytics-DELPHI	1-week-ahead	0.59	[0.28, 0.79]	P = 0.0235
CovidAnalytics-DELPHI	2-weeks-ahead	0.39	[0.02, 0.66]	P = 1
CovidAnalytics-DELPHI	3-weeks-ahead	0.44	[0.07, 0.70]	P = 0.492
CovidAnalytics-DELPHI	4-weeks-ahead	0.65	[0.36, 0.82]	P = 0.00489
Karlen-pypm	1-week-ahead	0.70	[0.45, 0.85]	P < 0.001
Karlen-pypm	2-weeks-ahead	0.59	[0.28, 0.79]	P = 0.0232
Karlen-pypm	3-weeks-ahead	0.18	[-0.20, 0.52]	P = 1
Karlen-pypm	4-weeks-ahead	0.39	[0.03, 0.67]	P = 0.907
RobertWalraven-ESG	1-week-ahead	0.73	[0.48, 0.87]	P < 0.001
RobertWalraven-ESG	2-weeks-ahead	0.72	[0.47, 0.86]	P < 0.001
RobertWalraven-ESG	3-weeks-ahead	0.75	[0.53, 0.88]	P < 0.001
RobertWalraven-ESG	4-weeks-ahead	0.79	[0.59, 0.90]	P < 0.001

Predictability versus 95% Prediction Interval (PI) Coverage

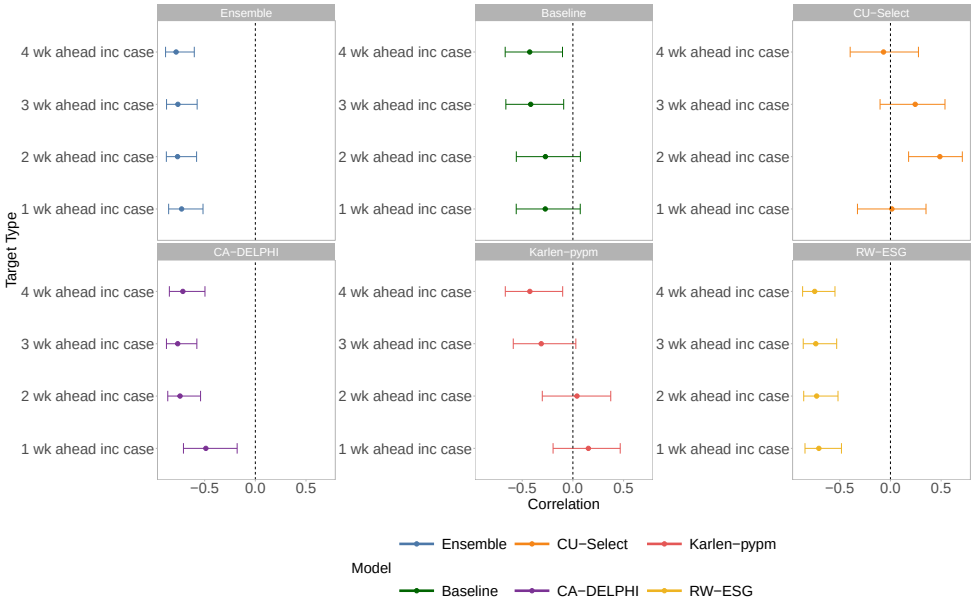


Figure SI G.42: **Pre-wave correlations between predictability and central 95% prediction interval (PI) coverage:** In Arizona before the onset of the studied wave (starting July 2021), we visualise the Pearson's correlations (Table SI G.4) between predictability and 95% PI coverage for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

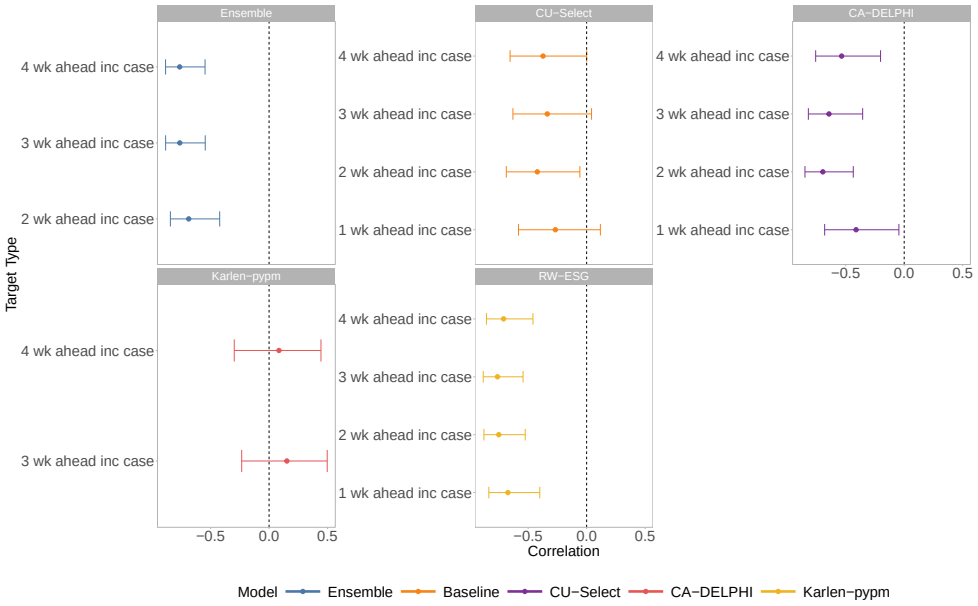


Figure SI G.43: **During-wave correlations between predictability and central 95% prediction interval (PI) coverage:** In Arizona during the studied wave (starting July 2021), we visualise the Pearson's correlations (Table SI G.4) between predictability and 95% PI coverage for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations. Note that no correlations are visualised for the one-week-ahead forecasts of the ensemble model, and for the two-weeks-ahead forecasts of the Karlen-pypm model, as these models had only one value within each trailing window for 95% PI coverage; 100%.

Table SI G.8: **Pre-wave correlations between predictability and central 95% Prediction Interval (PI) coverage:** In Arizona before the onset of the studied wave (starting July 2021), for each model and forecasted quantity, we report the Pearson's[586] correlation r between predictability and 95% PI, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	-0.73	[-0.86, -0.52]	P < 0.001
Ensemble	2-weeks-ahead	-0.77	[-0.88, -0.58]	P < 0.001
Ensemble	3-weeks-ahead	-0.76	[-0.88, -0.57]	P < 0.001
Ensemble	4-weeks-ahead	-0.78	[-0.89, -0.60]	P < 0.001
COVIDhub-baseline	1-week-ahead	-0.27	[-0.56, 0.07]	P = 1
COVIDhub-baseline	2-weeks-ahead	-0.27	[-0.56, 0.07]	P = 1
COVIDhub-baseline	3-weeks-ahead	-0.42	[-0.66, -0.09]	P = 0.344
COVIDhub-baseline	4-weeks-ahead	-0.43	[-0.67, -0.10]	P = 0.288
CU-select	1-week-ahead	0.01	[-0.33, 0.35]	P = 1
CU-select	2-weeks-ahead	0.49	[0.18, 0.71]	P = 0.0824
CU-select	3-weeks-ahead	0.24	[-0.10, 0.54]	P = 1
CU-select	4-weeks-ahead	-0.07	[-0.40, 0.28]	P = 1
CovidAnalytics-DELPHI	1-week-ahead	-0.49	[-0.71, -0.18]	P = 0.0821
CovidAnalytics-DELPHI	2-weeks-ahead	-0.74	[-0.86, -0.54]	P < 0.001
CovidAnalytics-DELPHI	3-weeks-ahead	-0.77	[-0.88, -0.58]	P < 0.001
CovidAnalytics-DELPHI	4-weeks-ahead	-0.71	[-0.85, -0.50]	P < 0.001
Karlen-pypm	1-week-ahead	0.15	[-0.20, 0.47]	P = 1
Karlen-pypm	2-weeks-ahead	0.04	[-0.30, 0.37]	P = 1
Karlen-pypm	3-weeks-ahead	-0.31	[-0.59, 0.03]	P = 1
Karlen-pypm	4-weeks-ahead	-0.42	[-0.67, -0.10]	P = 0.296
RobertWalraven-ESG	1-week-ahead	-0.71	[-0.84, -0.48]	P < 0.001
RobertWalraven-ESG	2-weeks-ahead	-0.73	[-0.86, -0.52]	P < 0.001
RobertWalraven-ESG	3-weeks-ahead	-0.74	[-0.86, -0.53]	P < 0.001
RobertWalraven-ESG	4-weeks-ahead	-0.75	[-0.87, -0.55]	P < 0.001

Table SI G.9: **During-wave correlations between predictability and central 95% Prediction Interval (PI) coverage:** In Arizona during the studied wave (starting July 2021), for each model and forecasted quantity, we report the Pearson's [586] correlation r between predictability and 95% PI, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	NA	NA	NA
Ensemble	2-weeks-ahead	-0.69	[-0.84, -0.42]	P < 0.001
Ensemble	3-weeks-ahead	-0.76	[-0.88, -0.55]	P < 0.001
Ensemble	4-weeks-ahead	-0.76	[-0.89, -0.55]	P < 0.001
COVIDhub-baseline	1-week-ahead	NA	NA	NA
COVIDhub-baseline	2-weeks-ahead	NA	NA	NA
COVIDhub-baseline	3-weeks-ahead	NA	NA	NA
COVIDhub-baseline	4-weeks-ahead	NA	NA	NA
CU-select	1-week-ahead	-0.27	[-0.58, 0.12]	P = 1
CU-select	2-weeks-ahead	-0.42	[-0.69, -0.06]	P = 0.432
CU-select	3-weeks-ahead	-0.34	[-0.63, 0.04]	P = 1
CU-select	4-weeks-ahead	-0.37	[-0.65, 0.00]	P = 0.865
CovidAnalytics-DELPHI	1-week-ahead	-0.41	[-0.68, -0.05]	P = 0.504
CovidAnalytics-DELPHI	2-weeks-ahead	-0.70	[-0.85, -0.44]	P < 0.001
CovidAnalytics-DELPHI	3-weeks-ahead	-0.64	[-0.82, -0.36]	P = 0.00381
CovidAnalytics-DELPHI	4-weeks-ahead	-0.53	[-0.76, -0.20]	P = 0.0572
Karlen-pypm	1-week-ahead	NA	NA	NA
Karlen-pypm	2-weeks-ahead	NA	NA	NA
Karlen-pypm	3-weeks-ahead	0.15	[-0.23, 0.50]	P = 1
Karlen-pypm	4-weeks-ahead	0.08	[-0.30, 0.44]	P = 1
RobertWalraven-ESG	1-week-ahead	-0.67	[-0.84, -0.40]	P = 0.00147
RobertWalraven-ESG	2-weeks-ahead	-0.75	[-0.88, -0.53]	P < 0.001
RobertWalraven-ESG	3-weeks-ahead	-0.76	[-0.88, -0.54]	P < 0.001
RobertWalraven-ESG	4-weeks-ahead	-0.71	[-0.86, -0.46]	P < 0.001

Predictability versus relative Weighted Interval Score (rWIS)

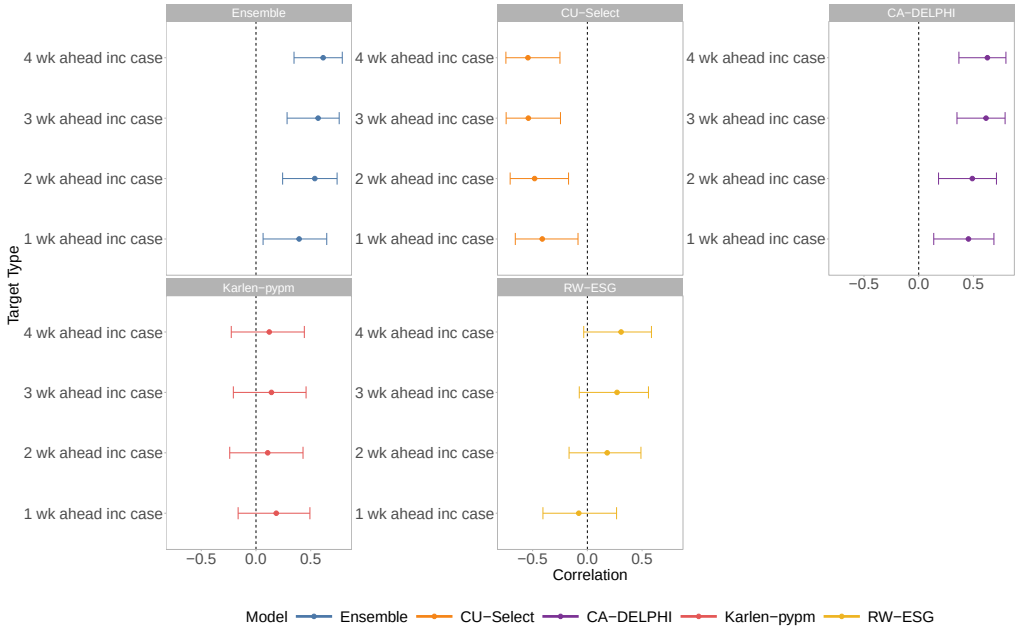


Figure SI G.44: **Pre-wave correlations between predictability and relative Weighted Interval Score (rWIS)** In Arizona before the onset of the studied wave (starting July 2021), we visualise the Pearson's correlations (Table SI G.5) between predictability and rWIS for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

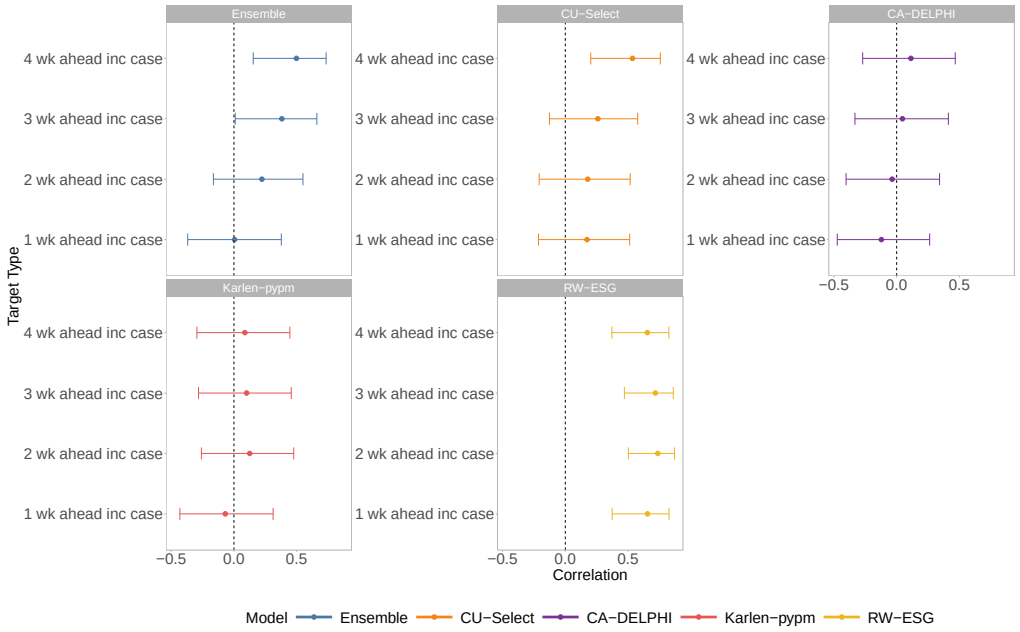


Figure SI G.45: **During-wave correlations between predictability and relative Weighted Interval Score (rWIS)** In Arizona before the onset of the studied wave (starting July 2021), we visualise the Pearson's correlations (Table SI G.5) between predictability and rWIS for each model and forecast target type, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations.

Table SI G.10: **Pre-wave correlations between predictability and relative Weighted Interval Score (rWIS)** In Arizona before the onset of the studied wave (starting July 2021), for each model and forecasted quantity, we report the Pearson's [586] correlation r between predictability and rWIS, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	0.39	[0.06, 0.65]	P = 0.421
Ensemble	2-weeks-ahead	0.54	[0.24, 0.74]	P = 0.021
Ensemble	3-weeks-ahead	0.57	[0.28, 0.76]	P = 0.00927
Ensemble	4-weeks-ahead	0.61	[0.35, 0.79]	P = 0.00229
COVIDhub-baseline	1-week-ahead	NA	NA	NA
COVIDhub-baseline	2-weeks-ahead	NA	NA	NA
COVIDhub-baseline	3-weeks-ahead	NA	NA	NA
COVIDhub-baseline	4-weeks-ahead	NA	NA	NA
CU-select	1-week-ahead	-0.41	[-0.66, -0.09]	P = 0.312
CU-select	2-weeks-ahead	-0.48	[-0.71, -0.17]	P = 0.078
CU-select	3-weeks-ahead	-0.54	[-0.74, -0.25]	P = 0.0201
CU-select	4-weeks-ahead	-0.54	[-0.74, -0.25]	P = 0.0184
CovidAnalytics-DELPHI	1-week-ahead	0.45	[0.14, 0.69]	P = 0.139
CovidAnalytics-DELPHI	2-weeks-ahead	0.49	[0.18, 0.71]	P = 0.0662
CovidAnalytics-DELPHI	3-weeks-ahead	0.61	[0.35, 0.79]	P = 0.00218
CovidAnalytics-DELPHI	4-weeks-ahead	0.63	[0.37, 0.80]	P = 0.00143
Karlen-pypm	1-week-ahead	0.19	[-0.16, 0.49]	P = 1
Karlen-pypm	2-weeks-ahead	0.11	[-0.24, 0.43]	P = 1
Karlen-pypm	3-weeks-ahead	0.14	[-0.21, 0.46]	P = 1
Karlen-pypm	4-weeks-ahead	0.12	[-0.23, 0.44]	P = 1
RobertWalraven-ESG	1-week-ahead	-0.08	[-0.41, 0.27]	P = 1
RobertWalraven-ESG	2-weeks-ahead	0.18	[-0.17, 0.49]	P = 1
RobertWalraven-ESG	3-weeks-ahead	0.27	[-0.07, 0.56]	P = 1
RobertWalraven-ESG	4-weeks-ahead	0.31	[-0.03, 0.59]	P = 1

Table SI G.11: **During-wave correlations between predictability and relative Weighted Interval Score (rWIS)** In Arizona during the studied wave (starting July 2021), for each model and forecasted quantity, we report the Pearson's[586] correlation r between predictability and rWIS, alongside the 95% Confidence Intervals (CIs) calculated using Fisher z-transformations and the Bonferroni-adjusted p-value for the two-sided hypothesis test with null hypothesis that $r = 0$.

Model	Target Type	r	95% CI	Bonferroni-adjusted p
Ensemble	1-week-ahead	0.01	[-0.37, 0.38]	P = 1
Ensemble	2-weeks-ahead	0.22	[-0.16, 0.55]	P = 1
Ensemble	3-weeks-ahead	0.38	[0.01, 0.66]	P = 0.897
Ensemble	4-weeks-ahead	0.50	[0.15, 0.73]	P = 0.139
COVIDhub-baseline	1-week-ahead	NA	NA	NA
COVIDhub-baseline	2-weeks-ahead	NA	NA	NA
COVIDhub-baseline	3-weeks-ahead	NA	NA	NA
COVIDhub-baseline	4-weeks-ahead	NA	NA	NA
CU-select	1-week-ahead	0.17	[-0.21, 0.51]	P = 1
CU-select	2-weeks-ahead	0.18	[-0.21, 0.52]	P = 1
CU-select	3-weeks-ahead	0.26	[-0.13, 0.58]	P = 1
CU-select	4-weeks-ahead	0.54	[0.20, 0.76]	P = 0.0665
CovidAnalytics-DELPHI	1-week-ahead	-0.12	[-0.47, 0.26]	P = 1
CovidAnalytics-DELPHI	2-weeks-ahead	-0.04	[-0.40, 0.34]	P = 1
CovidAnalytics-DELPHI	3-weeks-ahead	0.05	[-0.33, 0.41]	P = 1
CovidAnalytics-DELPHI	4-weeks-ahead	0.11	[-0.27, 0.47]	P = 1
Karlen-pypm	1-week-ahead	-0.07	[-0.43, 0.31]	P = 1
Karlen-pypm	2-weeks-ahead	0.13	[-0.26, 0.48]	P = 1
Karlen-pypm	3-weeks-ahead	0.10	[-0.28, 0.46]	P = 1
Karlen-pypm	4-weeks-ahead	0.09	[-0.30, 0.45]	P = 1
RobertWalraven-ESG	1-week-ahead	0.66	[0.37, 0.83]	P = 0.00308
RobertWalraven-ESG	2-weeks-ahead	0.74	[0.50, 0.87]	P < 0.001
RobertWalraven-ESG	3-weeks-ahead	0.72	[0.47, 0.86]	P < 0.001
RobertWalraven-ESG	4-weeks-ahead	0.65	[0.37, 0.83]	P = 0.00323

G.7 Distributionally robust optimisation for risk-averse decision-making

Expected tail loss (ETL) (also known as the *expected shortfall* or *conditional value at risk*) is a quantity central to the distributionally robust optimisation (DRO) literature[665,666]. ETL measures the risk associated to extreme events occurring at the upper tail of the loss distribution. Specifically, with Y denoting the outcome of interest (e.g., deaths, hospitalisations, or cases), ETL at level α is defined as*

$$\mathrm{ETL}_\alpha = \frac{1}{\alpha}\mathbb{E}\left[Y \cdot 1(Y \geq y_\alpha)\right],$$

where y_α is the $(1-\alpha)$-quantile of the distribution of Y (also known as the *value at risk* (VaR)). In words, ETL is the average loss incurred among the worst α% of outcomes. It is known that ETL and some other risk measures T of interest (such as the variance and mode of Y) are not *elicitable* from any proper scoring rule S, in the sense that there is no S for which predictions issued on the basis of T uniquely minimize the score[431,667,668].† In other words, evaluating forecasts based on S would yield conclusions inconsistent with the choice of T as a risk measure.

Arising within operations research and statistical learning, DRO is an alternative framework for decision-making under uncertainty. Above, we hinted at possible uncertainty in users' C/L ratios and in model rankings based on users' choice of evaluation trailing windows and spatial scale. Meanwhile, observed data during an epidemic are imperfect, impacted by time- and space-varying issues such as reporting errors, delayed reporting, and underrepresentation. All these issues can induce data distribution shifts and lead to metric-implied optimal decisions, from results on past data, no longer being optimal when making the decision in real-time on newly available data. Nevertheless, when seeking to make timely decisions (e.g., expanding ICU capacity) about important epidemic events (e.g., size of epidemic peak), a decision-maker would like the model-implied decision to be robust to uncertainty, distributional shifts, and worst-case scenarios.

We include DRO as a future tool for our framework to address these uncertainties in decision-making. Rather than assuming (unlikely) perfect knowledge of the outcome distribution Y or, more generally, of some relevant random quantity ξ indexing the loss function $h(x, \xi)$ associated to an action x, DRO attempts to minimise the worst-case cost of a decision across a range of plausible distributions of ξ. Mathematically, the DRO problem is

$$\inf_{x \in \mathcal{X}} \sup_{Q \in \mathcal{P}} \mathbb{E}^Q \left[h(x, \xi)\right], \tag{G.7.1}$$

where x is a decision variable in feasibility set $\mathcal{X}$ and $\mathcal{P}$ is an *ambiguity set* of probability distributions $\xi \sim Q$. A common choice of ambiguity set is an ϵ-neighborhood (with respect to some metric, discrepancy, or divergence, e.g., the Wasserstein metric or Kullback-Leibler divergence) around the empirical distribution Q_N (or a density estimate) of the observations $(\xi_1, \ldots, \xi_N)$.

*We assume for simplicity that Y has a continuous distribution function, in which case ETL takes a more intuitive form.

†However, it is known that the pair (VaR, ETL) is jointly elicitable in a certain sense[668].