Understanding mosquito vectors and methods for their control

Ben Lambert

Supervisors: Professor Sir Charles Godfray & Dr Ace North

A thesis presented for the degree of

Doctor of Philosophy

Somerville College
University of Oxford
UK
Trinity 2017
Acknowledgements

This work has been a labour of love for me although, for others, this is not an accurate description. This section is a meagre attempt to due justice to the patience and efforts of these people, in getting me from point A (a confused, slightly too old graduate student) to point B (a slightly less confused and, unfortunately, older graduate student).

First and foremost, I would like to thank my supervisors for the past three years, Ace North and Charles Godfray. Ace, I just checked my email history, and found 416 sent messages from me to you; to say you are a patient individual clearly does not suffice. Apart from answering my various queries through the years, I have learned much from you. I hope that my ineptitude has not put you off future supervisions, and that we can continue to work together in the future. Charles, thank you for our meetings. My scientific rigour and, I hope, the quality of my prose, have improved immeasurably since we met. Any good that comes from my work reflects on both of your efforts; any shortcomings reflects on my inability to listen to you.

There are various other people who have bettered my academic health throughout the years, including Armand Leroi, David Gavaghan, Astrid Iversen, everyone at Target Malaria, and Helen Byrne, and again, please accept my somewhat prosaic gratitude. I would like to especially thank Tom Churcher for including me in his work on near-infrared spectroscopy (on which Chapter 4 is based) after a chance meeting in Paris. Also, I would like to thank Tom for his excellent guidance, and for his contribution to this piece of work.

There are also those who have helped me, academically and sometimes less academically, throughout the years. Whilst this list is necessarily curtailed, I would like to thank Jean-michel Johnston, Hazel Tubman, Fergus Cooper, Joanna Raisbeck, Roman Walczak, and the various members of the Spartacan team.

There are two others who have done more than their fair share of listening to me over
these DPhilled years. Charlotte King, without you, I would not have got to Oxford in the first place. Claire Wheeler, without you, my life in Oxford would have not been as fun-filled, and truly great, as it was. Please accept my humble thanks, apologies, and future drinks, as an attempt to bring our trade imbalance further away from the red.

My parents, David Robinson and Vivien Lambert, and the rest of my family, I owe most to you. Throughout the years you have helped me considerably and, one day, I hope to be able to repay you. For now, you’ll have to accept these words as thanks.

Tatjana Detinova’s mass mosquito dissection experiments, consisting of tens of thousands of mosquitoes in Soviet Russia, Derek Charlwood’s seemingly global efforts, and the work of my experimental collaborators, Floyd Dowell and Maggy Sikulu-Lord, amongst other examples, make it clear that this work sits atop the broad shoulders of field entomologists past and present. I would like to dedicate this thesis to the arduous work of these insect people that do real science; I hope it does not disappoint.
The work of others

This section is an attempt to credit those people who directly contributed to the material within this thesis.

Ace North and Charles Godfray read through my introduction and provided useful comments throughout this process.

In Chapters 2 and 3, it should be recognised that Ace North and Charles Godfray met with me throughout the process of analysis, improving the content considerably. In Chapter 2, the original database of mark-release-recapture experiments was compiled by Guerra et al., 2014, and I thank them, and the work of field entomologists past. In Chapter 3, I compiled the database, but again owe much to the work of field entomologists in the past.

Tom Churcher was gracious enough to include me on his work on near-infrared spectroscopy, on which Chapter 4 is based. Many of the figures came about through our discussion and joint work on the manuscript, and I thank Tom for his efforts here. The near-infrared spectroscopy (NIRS) data was collected by my experimental collaborators, Floyd Dowell and Maggy Sikulu-Lord, and I thank them for their considerable work here.

Chapter 5, perhaps more than others, came from the joint work of myself and Ace North (Charles Godfray and Austin Burt also advised). My main contribution was the statistical model of rainfall, which we used to explore the effects of a homing endonuclease gene (HEG) release. Ace did much of the work of integrating my model into his spatial model of mosquito ecology (described in North, Burt, and Godfray, 2013). We both contributed equally to the running of simulations and figure generation.
A series of quotes illustrating malaria through the ages

...at the onset of his disease his temples bring fever to him, and afterwards they take the fever at sweat away...his are ‘blown’, his entrails...vertigo...his knees hurt him...[whatever] he eats and drinks, does not please him: that man is ill of the di’u illness.

– Medical description of a febrile illness called di’u, Babylonian-Assyrian text, circa 1900-600 BC

As one who has the shivering of the quartan so near,/ that he has his nails already pale/ and trembles all, still keeping the shade,/ such I became when those words were uttered.

– Discussion of quartan fevers (those appearing every four days) in The Inferno, Dante Alighieri, 1471

You are so very choleric of complexion./ Beware the mounting sun and all dejection./ Nor get yourself with sudden humours hot;/ For if you do, I dare well lay a groat/ That you shall have the tertian fevers pain,/ Or some ague that may well be your bane.

– Discussion of tertian fevers (those appearing every three days) in The Nun’s Priest’s Tale, Geoffrey Chaucer, c.1390s

... (he) hath got, as I take it, an ague ... hes in his fit now and does not talk after the wisest. He shall taste of my bottle: if he have never drunk wine afore it will go near to remove his fit ... Open your mouth: this will shake your shaking ... if all the wine in my bottle will recover him, I will help his ague.

– Discussion of ague (a term for an illness thought to be malaria in common use in England at the time) in The Tempest, William Shakespeare, c.1611
Beware beware, the Bight of Benin. One comes out where fifty went in.

– English jingle associated with slavery in West Africa, c.1800s

The poor white man, faint and weary

 Came to sit under our tree.

He has no mother to bring him milk;
No wife to grind his corn.
Let us pity the white man.

– An African song, c.1800s

Malaria is the great enemy of the explorer, the missionary, the planter, the merchant, the farmer, the soldier, the administrator, the villager and the poor...

– From The Prevention of Malaria, Ronald Ross, 1913

No one who has studied the epidemiology of malaria can fail to be impressed by the extreme diversity of the data recorded regarding parasite prevalence, period of transmission, degree of endemicity, epidemic potential and amenability to control measures in different regions of the world and even in different parts of the same country.

– From The Epidemiology and Control of Malaria, George Macdonald, 1957

Considering their impact, you might expect mosquitoes to get more attention than they do. Sharks kill fewer than a dozen people every year and in the U.S. they get a week dedicated to them on TV every year. Mosquitoes kill 50,000 times as many people, but if there’s a TV channel that features Mosquito Week, I haven’t heard about it.

– From a web article titled The Deadliest Animal in the World, Bill Gates, 2014
We are planning to have humans settling on Mars by 2023, which is 225 million miles away from earth. But malaria is just 0km from earth. If we can solve problems at 225 million miles away from us, then there is no reason that we can't defeat malaria here, in our living room.

– From the Target Malaria presentation at the Grand Challenges meeting, Abdoulaye Diabate, 2016
Contents

1 Introduction ......................................................... 15
  1.1 Abstract .......................................................... 15
  1.2 Introduction ...................................................... 17
  1.3 The largely forgotten impact of mosquito borne disease in the developed
       world throughout history ........................................ 26
  1.4 The impact of mosquito borne disease in the developing world today .......... 37
  1.5 Mosquitoes as vectors ............................................ 49
  1.6 Disease control since Ronald Ross ................................ 52
  1.7 Why we need to understand mosquito vectors .......................... 64
  1.8 The importance of mosquito senescence ............................ 72
  1.9 Future vector control by genetic drive ............................ 75
  1.10 Motivating this thesis ............................................ 81

2 A meta-analysis of mark-release-recapture experiments of mosquitoes ...... 83
  2.1 Abstract .......................................................... 83
  2.2 Introduction ...................................................... 84
  2.3 Materials and methods ............................................ 87
    2.3.1 Data .......................................................... 87
CONTENTS

2.3.2 Statistical model ......................................................... 87
2.3.3 Individual time-series estimates ..................................... 91
2.3.4 Estimating lifespan at the species, genus and overall groupings . . 92
2.3.5 Testing for age-dependent mortality in wild mosquitoes ............ 95
2.3.6 Model estimation by MCMC ............................................. 95
2.3.7 K-Fold cross validation .................................................. 98

2.4 Results ........................................................................ 102
2.4.1 Mosquito lifespan varies across species although variation is less marked at the genus level .............................................. 102
2.4.2 A small proportion of mosquitoes are likely responsible for the bulk burden of disease for malaria and Zika ......................... 104
2.4.3 The longevity of mosquitoes does not vary systematically over the range of temperatures of MRR study locations ...................... 106
2.4.4 Male mosquitoes have shorter lives than females ..................... 106
2.4.5 Blood- and sugar-fed mosquitoes do not live longer than those that were not fed before release ..................................... 107
2.4.6 Wild mosquitoes do not appear to experience strong age-dependent mortality ....................................................... 107

2.5 Discussion .................................................................... 119
2.5.1 The statistical power of MRR studies ................................. 122
2.5.2 MRR design ................................................................. 124

2.6 Conclusion .................................................................... 127

3 A meta-analysis of mosquito dissection experiments .................. 129
3.1 Abstract ................................................................... 129
3.2 Introduction ................................................................. 130
3.3 Method ................................................................. 138
  3.3.1 Collection of dissection data ................................ 138
  3.3.2 Statistical analysis of dissection data ................... 141
  3.3.3 Data collection of gonotrophic cycle duration ........... 149
  3.3.4 Statistical analysis of gonotrophic cycle data .......... 150
  3.3.5 Conversion of lifespan from physiological to calendar age 152
3.4 Results ........................................................... 158
  3.4.1 Wild mosquito lifespan in gonotrophic cycles .......... 158
  3.4.2 Gonotrophic cycle duration ............................... 159
  3.4.3 Wild mosquito lifespan in calendar age .................. 160
  3.4.4 The evidence for age-dependent mortality ............... 160
3.5 Discussion ....................................................... 167
  3.5.1 Comparison with MRR estimates .......................... 167
  3.5.2 Comparison with laboratory-based methods ............... 172
  3.5.3 Does age-dependence matter? ............................. 174
3.6 Conclusion ....................................................... 175

4 Using near-infrared spectroscopy to estimate mosquito population mean lifespan 179
  4.1 Abstract .......................................................... 179
  4.2 Introduction ..................................................... 180
  4.3 Method .......................................................... 185
    4.3.1 Data ......................................................... 185
    4.3.2 Machine learning method: existent work ............... 186
    4.3.3 Machine learning method: our approach ............... 189
    4.3.4 Cross validation and bias correction .................... 190
4.3.5 Surrogate Bayesian model ........................................... 193
4.3.6 Population lifespan inference ................................. 194
4.4 Results ................................................................. 196
  4.4.1 Accurate prediction of individual mosquito age within a given experiment ........................................... 196
  4.4.2 Individual prediction errors decrease with sample size ................................. 198
  4.4.3 Inferring population mean lifespan is possible with modest sample sizes .................... 199
  4.4.4 The bulk of uncertainty in population lifespan estimates is due to sampling variability ........................................... 201
  4.4.5 Between-study prediction is poor ........................................... 203
4.5 Discussion ............................................................. 207
4.6 Conclusion ............................................................. 210

5 Using a homing endonuclease to control vector populations in a landscape with temporal heterogeneity ........................................... 213
  5.1 Abstract ............................................................. 213
  5.2 Introduction ............................................................. 214
  5.3 Method ............................................................... 216
    5.3.1 Mosquito population model ........................................... 216
    5.3.2 Linking rainfall with aquatic habitat density ........................................... 220
    5.3.3 A general model for rainfall ........................................... 221
    5.3.4 HEG releases ............................................................. 222
  5.4 Results ............................................................... 223
    5.4.1 Constant environment ........................................... 223
    5.4.2 HEG dynamics are more variable in a seasonal environment ........................................... 224
    5.4.3 The start of the wet season is the optimal time for release ........................................... 225
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4.4</td>
<td>In seasonal environments eradication is easier to achieve if the average aquatic habitat density is low or high rather than intermediate.</td>
<td>225</td>
</tr>
<tr>
<td>5.4.5</td>
<td>Random variation in rainfall generally increases the probability that a HEG will induce population extinction.</td>
<td>226</td>
</tr>
<tr>
<td>5.4.6</td>
<td>Releases during the first half of the wet season result in the highest chance of population extinction, although in some cases the time taken for extinction to occur may be lower by releasing HEGs during the dry season.</td>
<td>227</td>
</tr>
<tr>
<td>5.5</td>
<td>Discussion</td>
<td>237</td>
</tr>
<tr>
<td>5.6</td>
<td>Conclusion</td>
<td>242</td>
</tr>
<tr>
<td>6</td>
<td>Conclusion</td>
<td>243</td>
</tr>
<tr>
<td>7</td>
<td>Appendix</td>
<td>257</td>
</tr>
<tr>
<td>7.1</td>
<td>A mathematical model of senescence</td>
<td>257</td>
</tr>
<tr>
<td>7.2</td>
<td>Studies included in the MRR meta-analysis</td>
<td>271</td>
</tr>
<tr>
<td>7.3</td>
<td>Studies included in the dissection study meta-analysis</td>
<td>272</td>
</tr>
<tr>
<td>7.4</td>
<td>Studies included in the gonotrophic cycle duration meta-analysis</td>
<td>273</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Abstract

Mosquitoes spread diseases that shorten and worsen the lives of many people, chiefly children in poor countries, around the world. Since Ronald Ross’ discovery at the end of the nineteenth century that mosquitoes transmit malaria, field entomologists have collected a great deal of information about mosquito ecology. Despite this tremendous effort, there still remain significant gaps in our knowledge of mosquito ecology, in part, reflecting the significant variation in mosquito ecology across species and geographies. The main aim of this thesis is an attempt to synthesise the substantial information that field entomologists have collected on mosquito lifespan. In Chapters 2 and 3, I conduct meta-analyses of the two predominant approaches used to estimate mosquito lifespan: mark-release-recapture experiments and female mosquito dissection-based studies, respectively. These analyses produce estimates of mosquito lifespan by species and genus, and more broadly, allow for an appraisal of these two experimental approaches. In Chapter 4, I describe a recently developed approach, known as near-infrared spectroscopy, which enables estimation of the
age of individual mosquitoes, and then perform an in silico analysis to explore the use of
this technology for estimating the average lifespan of wild populations of mosquitoes.

The emergence of mosquito resistance to the main insecticides used in vector control,
along with the concerning recent discovery that the malaria parasites in Asia are becoming
resistant to artemisinin - an important drug used to treat malaria - highlight the need
for novel approaches to control disease transmission. Some recently-proposed approaches
involve genetic modification of the mosquito vectors, for example, to render them incapable
of acting as hosts for disease or to reduce their fecundity. In Chapter 5, I model the impact
of a release of mosquitoes carrying a genetic construct known as a homing endonuclease,
which has been constructed to bias the sex of mosquito offspring towards males, in compu-
tational environments that capture some facets of the real life landscapes where mosquito
borne disease is rife.

About a century ago, the famous Italian Malariologist Giovanni Grassi declared that
malaria was a “giant with clay feet”; reflecting the optimism, in some academic circles at
the time, that eradication of this terrible disease would soon occur. Unfortunately, a cen-
tury of often unsuccessful attempts to control and eradicate malaria, and other mosquito
borne diseases, would follow Grassi’s statement, meaning that this fight is likely to con-
tinue throughout the twenty-first century. We now know much more about mosquitoes and
mosquito borne disease than we did a generation ago, but there is still crucial information
that we do not. In this thesis, I argue that in order to make significant inroads to disease
eradication, further research on mosquito ecology is crucial. Only when we better under-
stand our unwitting mosquito foe, can we design and implement effective disease control
measures that are so desperately needed in those most desperate parts of the world.
In terms of the number of people they sicken and kill, mosquitoes are more dangerous than any other organism on earth, including humans (Gates, 2014). At the peak of their breeding season, mosquitoes outnumber every other animal on the planet, except ants and termites (Gates, 2014). Malaria is the most deadly of the mosquito borne diseases, which is vectored by anopheline mosquitoes and caused by microorganisms of the genus *Plasmodium*. Malaria has plagued humankind throughout our history: the 4700 year old Chinese tome *The Nei Ching*, probably the world’s oldest medical textbook, refers to conditions characterised by recurrent fevers and spleen enlargement, typical of malaria (Shampo and Kyle, 1989); it has been argued that Alexander the Great died of *P. falciparum* malaria on route to India (Carter and Mendis, 2002); Shakespeare mentions ague (the English word for the disease in common usage until the 19th Century) in eight of his plays (Reiter, 2000). Whilst likely exaggerated, it has been suggested that malaria accounts for half of all human deaths through history (Whitfield, 2002). In the twentieth century alone, it has been estimated that malaria killed 300 million people (Carter and Mendis, 2002), as much as five times the combined death toll from both world wars (Britannica, 2017a; Britannica, 2017b). Regardless of its historical importance, malaria today sickens around 300 million and likely kills between 0.5 and 1.2 million people each year (World Health Organisation, 2015; Murray et al., 2012), most of whom are children born into poverty in the tropics. Mosquitoes also vector other diseases that continue to inflict devastating human suffering worldwide. These include filariasis - a disease caused by macroscopic worms that often causes severe disfigurement in those it affects - which has an annual incidence of around 40 million cases worldwide (World Health Organisation, 2017a). *Aedes* mosquitoes vector a febrile disease called chikungunya whose global spread is widening (Charrel, 2007), having recently caused outbreaks in Italy (Rezza et al., 2007). Chikungunya, like many other
mosquito borne diseases, has no cure. The ecological disruption caused by World War II expanded the geographic distribution of *Ae. aegypti* (Gubler, 2002), which are major vectors of chikungunya, as well as the deadly dengue virus. Troop movements during this period likely accelerated the spread of dengue virus throughout Asia and, today, it has been estimated that there are 390 million dengue infections each year (Bhatt et al., 2013) and that 3.9 billion people - over half the world’s population - are at risk of contracting this disease (Brady et al., 2012). Whilst dengue infection is often asymptomatic, it can result in influenza-like symptoms known as *dengue fever* and also a deadly condition referred to as *dengue hemorrhagic fever* (Guha-Sapir and Schimmer, 2005), where affected patients bleed from their nose and mouth before falling into a state of shock from which they may never recover. In February 2016, the World Health Organisation declared a ‘Public Health Emergency of International Concern’, following the rapid spread of the Zika virus (mainly vectored by *Ae. aegypti* and *Ae. albopictus* mosquitoes) through Latin America (World Health Organisation, 2016b). The prevailing opinion is that Zika infection is behind the tragic rise in cases of microcephaly in newborn children in Brazil observed recently (Mlakar et al., 2016; Oliveira Melo et al., 2016).

Set against this backdrop of illness and suffering - both historical and present - it is surprising that mosquitoes do not get more media attention than they do. Indeed, as a researcher in the field, it is not uncommon to encounter views from members of the public in the developed world, to the tune of ‘I thought we had solved malaria, so why are you studying mosquitoes?’ . Academically, there has been over one hundred years of research into the organism, initially catalysed by Ronald Ross’ discovery in 1897 that anopheline mosquitoes could transmit the malarial parasite, and implicated mosquitoes in the transmission of human malaria (Cox, 2010). Throughout the twentieth century, the pendulum has swung frequently between the two predominant approaches used to
control malaria (paralleled in other mosquito borne diseases): targeting the human stages of the disease using drugs (for malaria, quinine then chloroquine then artemisinin) and vector control methods (Packard, 2007). In the 1980s, advances in our understanding of immunology and vaccination technology lead to a shift towards interventions that targeted the human stages of vector borne disease (Godfray, 2013). Indeed, in their landmark text on the mathematical epidemiology of infectious diseases Anderson and May, 1992 declared, ‘Today, with the intense interest in the development of vaccines and the application of new molecular and biochemical techniques to the study of acquired immunity in humans, the research emphasis has changed from entomology to infections in humans.’ More recently, the popularity of vector control methods has been resurgent (and largely responsible for the recent gains against malaria in Sub-Saharan Africa), partly due to the difficulty in developing vaccines and drugs that can effectively target pathogenesis across the diverse life stages of malarial and filarial parasites in humans (Greenwood et al., 2008). In Cambodia in 2009, forms of the malarial parasite resistant to artemisinin (the main chemical compound from which drugs used to treat human malaria are derived) were reported (Dondorp et al., 2009). These resistant forms of the parasite have subsequently spread throughout Asia (Ashley et al., 2014) and now its potential to invade Africa threatens to inflict as much damage to human life as resistance to chloroquine caused 50 years ago (White, 2010). These recently emergent parasitological hurdles underscore the continuing importance of vector control in the twenty-first century and have, in part, lead to calls for a generation of novel approaches to vector control.

To design the most effective vector control interventions, it is essential to have a detailed knowledge of mosquito ecology. The transmission of vector borne disease is particularly sensitive to the amount of time that adult mosquitoes live in the wild (Ross, 1910; Macdonald, 1957; Smith, Dushoff, and McKenzie, 2004). Mosquitoes that live longer bite more humans
and, hence, have a greater chance to become infected with the aetiological agent, survive its extrinsic incubation period, and transmit the disease. Whilst it is fairly straightforward to estimate the mean lifespan of mosquitoes laboratory colonies, this task is nontrivial for wild mosquitoes. Historically, two approaches have been used to estimate the lifespan of wild mosquitoes. The first is mark-release-recapture experiments (see, for example, Silver, 2007) where a number of lab-reared or wild-caught mosquitoes are marked, typically with a fluorescent die, then released into the wild. By monitoring the number of marked mosquitoes that are recaptured over time, it is possible to estimate the combined rate of mosquito mortality and dispersal (if the number recaptured rapidly decays, mosquitoes die and fly away from the study area at a correspondingly fast rate). The second approach is based on the dissection of female mosquitoes. Female mosquitoes undergo changes in their reproductive anatomy that occur at roughly uniform intervals throughout their lifetime, providing a biological ‘clock’ which can be used to assess the physiological age of individual specimens by dissection (Polovodova, 1949; Detinova, 1962). By determining the age of a large number of specimens this allows estimation of mosquito lifespan. Whilst both of these approaches provide valuable (and complimentary) information about mosquito ecology, the individual experiments are costly and require large numbers of mosquitoes to produce accurate estimates of mosquito lifespan. Further, there is considerable variation in the estimates of mosquito lifespan produced across studies, presumably due to differences in experimental protocol, the particular mosquito species studied, and the climate and geography of the study region (Guerra et al., 2014). In Chapters 2 and 3, I present the results of analyses of data from a large number of mark-release-recapture (MRR) experiments and dissection-based studies, respectively. These meta-analyses synthesise the results of the arduous work done by field entomologists and provide robust estimates of wild mosquito lifespan. A benefit of analysing the results of many studies is that this allows us (my col-
1.2. INTRODUCTION

To our knowledge, this is the first attempt to do this, and we hope this will prove useful for researchers and public health officials who study the epidemiology of various mosquito borne diseases. The simultaneous analysis of data from the two types of experiments also allows us to appraise both approaches in a way that was previously not possible.

Some of the most effective methods of vector control, including insecticide treated nets and indoor residual spraying, aim to kill adult mosquitoes. At a population level, these approaches should lead to a reduction in the average lifespan of mosquitoes, resulting in lower disease transmission. In the past, the main way to assess vector control measures at an operational level was by surveying the percentage of the human population protected by interventions. However, resistance to pyrethroids - the main insecticide used in mosquito vector control and the only class used for treating bednets - is now widespread throughout Africa (Ranson et al., 2011; Basilua Kanza et al., 2012; Mulamba et al., 2014; Ranson and Lissenden, 2016), likely weakening the effect of traditional vector control methods. This resistance also means that intervention coverage is no longer a strong indicator of vector control. As I discussed above, MRR experiments and mosquito dissection studies can be used to estimate the lifespan of wild mosquito populations. Whilst these methods are currently the main approaches used to study many aspects of mosquito ecology, these experiments are time consuming, costly, and require specialist knowledge of mosquito biology to be effectively implemented. In Chapter 4, I discuss a novel approach to estimating mosquito lifespan using near-infrared spectroscopy (NIRS).

As a mosquito ages, it undergoes changes in its body chemistry. These changes can be detected by shining near-infrared light through the head and thorax of mosquito specimens and examining the resultant spectrum. Further, previous work using laboratory mosquitoes has demonstrated that the changes to the near-infrared spectrum throughout a mosquito’s
lifetime are regular enough to be able to predict the age of individual specimens (Sikulu et al., 2010; Sikulu et al., 2011; Dowell, Noutcha, and Michel, 2011; Sikulu et al., 2014; Liebman et al., 2015). In Chapter 4, I present an analysis of a dataset which we compiled from a number of previously-published NIRS experiments where, for each sample, we have the near-infrared spectrum of a mosquito of known age. In the previously-published work, the statistical models used to predict the age of mosquitoes have been relatively basic. In this chapter, I apply more recently-developed machine learning models to the task of predicting the age of individual specimens in our database, resulting in a greater predictive accuracy than has previously been reported. Since our database consists of various previous studies, this also allows us, for the first time, to determine how heterogeneity in study protocol impacts the predictive accuracy of NIRS.

To gauge the effectiveness of vector control methods that aim to kill adult mosquitoes, we wish to know how the mean lifespan of the mosquito population is reduced as the interventions are implemented. This focus on the population level has been neglected by previous NIRS studies, which tend to focus on predicting the age of individual specimens. In Chapter 4, I use in silico populations of mosquitoes of a known age distribution, to demonstrate that NIRS can be used to infer changes in the population mean age across those age ranges likely to be encountered in applied vector control. Since the cost of NIRS is mostly upfront (the initial cost of the machine) with it being relatively easy to scan individual mosquito samples, I argue that this approach may provide a viable means of surveying mosquito vector populations in the future.

Since the turn of the twenty-first century, there has been unprecedented reduction in the human costs of malaria driven largely by increasing coverage of insecticide treated nets (Bhatt et al., 2015). It has been estimated that 663 million disease cases may have been averted due to the combination of insecticide treated nets, indoor residual spraying
and artemisinin-based combination therapy from 2000-2015 (Bhatt et al., 2015). The increasingly widespread resistance of mosquitoes to insecticides, climate change, along with the emergent threat of malaria parasite resistance, however, threatens to halt and, perhaps, even reverse these trends. Worryingly, in 2016, progress against malaria appears to have stalled, with an increase of around 5 million cases over 2015 estimated (World Health Organisation, 2017c). This hints that the future gains against malaria may be harder to attain and suggests that new strategies are required to continue to alleviate the disease burden in the future.

Endonucleases are genes that code for proteins that cleave DNA at specific recognition sites in the genome and occur in a variety of organisms. A class of these genes known as homing endonucleases (HEGs) are thought to be ‘selfish genes’, which use a trick to rapidly spread through populations. In heterozygotes for this gene (where one chromosome carries the gene but the other does not), a HEG protein cuts the opposite chromosome and the cell sets about repairing it. The cell’s repair mechanism, however, uses the uncut chromosome - the one carrying the HEG - as a blueprint for the repair operation. From the HEG’s perspective, this is fortunate because it is embedded in the blueprint, meaning it will be copied over to the other chromosome and, in doing so, convert the heterozygote into a homozygote. This huge selective advantage means that a newly-introduced HEG will spread from rare to fixation in a population (Goddard and Burt, 1999). Burt, 2003 first argued that HEGs could be used to control mosquito vector populations. Burt imagined that there were a number of ways that this could be done. For example, if the HEG recognised a gene essential for the pathogen but of low importance for the mosquito host, this could prevent the pathogen’s transmission, halting the spread of disease. Alternatively, if the HEG cut a site that was crucial for mosquito survival or reproduction, then the spread of this genetic element could knock down the population density of mosquitoes to
levels where the stochastic loss of the parasite is possible. In Chapter 5, I present the results from computer simulations of a different sort of HEG; again originally proposed by Burt, 2003: suppose a HEG was inserted onto the Y chromosome but was engineered to recognise a site on the X chromosome (and was expressed at meiosis), then only the Y gametes would be viable, resulting in entirely male offspring (mosquitoes have the same sex determination mechanism as humans). These so-called ‘Y drive’ genes would spread throughout the population and, in doing so, increasingly skew the population sex ratio towards males. Much like a party all-too-typical of human adolescence, eventually the few remaining females would be stochastically lost and the population would be locally eliminated.

In Chapter 5, I build on previous modelling work of Y-drive HEGs (Burt, 2003; Dere-dec, Burt, and Godfray, 2008; Deredec, Godfray, and Burt, 2011; North, Burt, and Godfray, 2013) to consider its spread in more realistic models of the underlying landscape. Specifically, I extend a spatial model introduced by North, Burt, and Godfray, 2013 that incorporates key aspects of mosquito ecology. North, Burt, and Godfray, 2013 considers how spatial characteristics of a landscape influence the impact of a Y-drive HEG on local mosquito populations. Here, I use computational simulation to investigate how temporal characteristics of a landscape - the availability of mosquito aquatic habitats, driven by short-term fluctuations in rainfall - affect the use of such a HEG to control vector populations. Encouragingly, the results that I present indicate that if a Y-drive HEG were deployed at the optimal time of year (also determined through simulation) then the resultant intervention may actually be bolstered by temporal heterogeneity. I conclude this chapter by using computational simulation to investigate the spread of a Y-drive HEG in two case study locations: one in Kenya, in tropical east Africa, the other in Mali, in the Sahel region in West Africa, and demonstrate that a HEG may an effective tool for vector
1.2. INTRODUCTION

In Chapter 1, I provide a historical narrative of the impact of vector borne disease on humans and the methods used to control it. This torrid history highlights the paramount importance of these diseases and conveys the practical difficulty of eradicating these scourges of humanity. I also believe that this historical perspective suggests that there is likely to be no silver bullet for controlling, and later eradicating, mosquito borne diseases. I do argue, however, that vector control methods are a crucial part of those strategies most likely to put an end to such manifold human suffering.

The role of the mosquito in human malaria transmission was first determined by Ronald Ross and other researchers at the turn of the twentieth century. In this chapter, I discuss the Ross-Macdonald mathematical model of malaria that underlies the main approaches to vector control and use it to highlight the need to better understand mosquitoes; specifically, the importance of knowing how long mosquitoes live. I also discuss how the nature of mosquito mortality matters. Historically, it has been assumed that mosquitoes, unlike humans, very rarely reach ages where they experience rapid physiological decline leading to death. If this is incorrect, mathematical models have hinted that we need to re-evaluate our current approaches to vector control. Whilst, in Chapters 3 and 4 I present evidence that mosquito senility is not a strong factor in wild mosquito mortality, in this introductory chapter I describe the previously-published attempts to look for senescence and the mathematical models that motivate this hunt.

The remaining part of this chapter is a short survey of gene-drive mechanisms that have been proposed to control mosquito populations. This area is expansive and I cannot do it full justice here. Instead, I focus on the mathematical models that have been used to investigate the performance of a few genetic constructs that are currently being considered for vector control.
1.3 The largely forgotten impact of mosquito borne disease in the developed world throughout history

When unable to defend herself by the sword Rome could defend herself by means of the fever.

– The poet Godfrey of Viterbo, 1167

Ancient Rome was frequently paralysed by outbreaks of malaria (Celli, 1925). Being millennia before Louis Pasteur and Robert Koch provided the final proof of the germ theory of disease, the Romans turned to celestial origins. A temple to Dea Febris - the goddess of fever (see Figure 1.1) - was constructed on the Palatine hill, possibly the site of the original settlement of Rome by Romulus. According to Valerius Maximus (an author of historical anecdotes living during the reign of Tiberius), the worship of the goddess of fever was driven by fear, “When they worship other deities, they expect to receive a benefit. However fever is worshipped, so that she will cause less harm...” (Maximus and Valerius, 1888). The terror of malaria was so widespread that there even existed a febris cult devoted to the deity. Lest we imagine that belief in the divine providence of malaria only existed during Classical antiquity, the febris cult was active in Italy well into the twentieth century (Sallares, 2002).

The historical presence of malaria has also affected our language. Interestingly, the word abracadabra, ubiquitous in children’s pantomimes, may have originated as a Roman treatment for semitertian fever. Quintus Serenus Sammonicus, physician to the Roman emperor Caracalla, dictated that patients suffering fever wear an amulet inscribed with an incantation based on the word abracadabra for nine days (see Figure 1.2; Serenus and Pépin, 1950; Hempelmann and Krafts, 2013). Whilst some scholars believe that abracadabra has no meaning, others believe it may have been derived from the Hebrew words Abrai seda
1.3. MOSQUITO-BORNE DISEASE IN DEVELOPED WORLD

Figure 1.1: A statue of Dea Febris, goddess of fever, by unknown artist. Reproduced from Tanogabo, 2015.

brai meaning “Out, bad spirit, out” or, alternatively, from the Hebrew phrase Ab, Ruach, Dabar meaning “Father, Holy Ghost, Word” (Wootton, 1910).

Human efforts to transform the land for agriculture in the 1600s abetted the northwards expansion of malaria from southern Europe, where it was present from at least the 4th century BC (Packard, 2007). The seasonality and cooler temperatures likely thwarted *P. falciparum* from establishing, but *P. vivax* and *P. malariae* could and did spread into England and as far north as Russia (Packard, 2007). During seventeenth century England, the expanding population of well-off folk in London and other cities demanded more food. This rising demand for agricultural produce lead to efforts to drain the marshlands of southeast England and the Fenlands of Cambridgeshire and Lincolnshire to make way for farmland. These drainage schemes created the ideal habitats for anopheline mosquitoes. In attempts to hold back the sea, embankments were constructed that altered the flow of water over the land and lead to the development of brackish pools of water and swamps.
A. atroparvus is adapted to mildly saline aquatic habitats and was probably the primary vector of malaria in these parts of England (Reiter, 2000).

Malaria became a leading cause of death in the marsh areas of mostly southeast and eastern England in the sixteenth and seventeenth century (Packard, 2007). Malaria was so widespread in the marshlands that it is likely this is the reason that crude burial rates were so much higher in marshland areas than in other parts of the country (see Figure 1.3; Dobson, 1980). (Recently, however, some have argued that the misattribution of ague symptoms to malaria, and low mortality risk of *P. vivax* - the type of malarial parasite likely present in England at the time - means that malaria was perhaps not the main driver of high coastal death rates; Hutchinson and Lindsay, 2006.) The inflated death rate in the southeast was also probably fuelled by the migration of people of the “lesser sort” - people often of undesirable heritage, including smugglers and shepherds - to the region, who were willing to risk their lives for a chance at economic betterment (Dobson, 2003).
Whilst this migration may have benefited the local economy, the large influx of individuals lacking any immunity to malaria could have worsened the marshland malarial situation (Packard, 2007). It is unfamiliar for us to associate what is today the prime commuter belt for London with malaria, but in the sixteenth till the late nineteenth century this was conceived wisdom,

*In Kent, and all along on Essex side A troupe of cruel fevers did reside: Of severall Country agues lay an Hoast. And most of them, who had this place forsooke Were either slaine by them, or prisoners tooke.*

– The poet George Wither, 1588-1667

One of the strongest pieces of evidence that ague (the common term for the condition causing intermittent fevers in England at the time) was indeed malaria, is the identity of its cure.

Peruvian Indians chewed the bark of the cinchona tree - possibly to avoid shivering in Spanish-owned mines because of quinine’s muscle relaxant effects - but it is not believed that they did this to treat malaria (Meshnick and Dobson, 2001). The countess of Chinchon and her husband are falsely credited with bringing the bark back to Europe, and it is after her that Linneas in 1742 named the tree, cinchona, although the more common name for the powder used to treat ague was Jesuits’ powder (Jesuit priests were the first to bring cinchona bark from the New World) or Peruvian bark (Bruce-Chwatt, 1988).

The active ingredient of cinchona bark is quinine, possibly the drug responsible for alleviating more suffering in human history than any other single chemical. Quinine is still used to treat malaria today, but most people do not know that its clinical treatment was pioneered on patients suffering from ague in the salt marshes of Essex. Robert Talbor - a relatively uneducated man - developed an effective treatment regimen for ague that he published in 1672 (Reiter, 2000). His concoction, perhaps the first patent medicine,
was essentially a mix of white wine and cinchona powder. The success of his treatment, however, was enough to propel him to fame and fortune, and he was knighted in 1678 after he cured the King Charles II of ague.

The increasing cattle population between 1840 and 1910 provided anopheline vectors with another source of blood and, together with a decreasing acreage of marshland, contributed to a major decline in malaria over the period (Kuhn et al., 2003; Packard, 2007). Indeed, knowledge that malaria once plagued the British Isles has largely been forgotten. This is not without reason - whilst more than 52,000 cases of malaria have been imported since 1953, none of these have lead to secondary cases (Kuhn et al., 2003).

Whilst there is evidence that malaria has likely been with us throughout recorded history, the parasite itself, and its association with humans, predates this. It is believed that the ancestors to the malarial parasites first became adapted to the larvae of an aquatic insect of the order Diptera, to which mosquitoes and other blood-sucking flies belong, which first appeared around 150 million to 200 million years ago (Carter and Mendis, 2002). Either during this era or following it, the ancestral malarial parasites became adapted to the two-host life cycle that characterises them now (Carter and Mendis, 2002). Since the emergence of this first malarial parasite, adaptation and evolution of the parasite caused the phylogeny to radiate into many different lines (see Figure 1.4). The malarial parasites came to parasitise many different animal species, of which the greatest number have been recorded amongst primates (Garnham, 1966). Of those that affect humans, there are five, all of the genus *Plasmodium*: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*. The five human malarial parasites have quite different evolutionary histories, and some of these are more related to animal malaria species than they are to one another. The most distinct of the parasites is *P. falciparum*, which today causes the most serious forms of illness. *P. falciparum* is closely related to *P. reichenowi* (Escalante and Ayala,
1.3. MOSQUITO-BORNE DISEASE IN DEVELOPED WORLD

1994; Escalante, Barrio, and Ayala, 1995; Pacheco et al., 2013), which causes malaria in chimpanzees and, interestingly, each of these are more related to *P. gallinaceum*, the aetiological agent of avian malaria, than they are to many other forms that target mammals (Escalante and Ayala, 1995). The most recent common ancestor of *P. falciparum*, *P. reichenowi* and the avian malarial parasite is estimated to have existed around 130 million years ago (Escalante and Ayala, 1995), with the divergence between *P. falciparum* and *P. reichenowi* having occurred around 4 million to 10 million years ago (Escalante and Ayala, 1994; Escalante, Barrio, and Ayala, 1995). This latter period overlaps with the period, around 5 million years ago, that the human line diverged from the African great apes (Kumar and Hedges, 1998). The other four forms of the parasite exist in clades which include other mammalian parasites (Escalante, Barrio, and Ayala, 1995; Singh et al., 2004; Pacheco et al., 2013). It has been suggested that all forms of human malaria have African sources and that their characteristics (including the illnesses they inflict) have evolved over the past 100,000 years, governed by climatic events and human migration (Carter and Mendis, 2002). Whilst opinions differ on exactly where and how different forms of malaria developed, it is clear that malaria is an ancient human affliction, and that us and the parasite likely co-evolved throughout our joint history.

Malaria is sadly not the only disease borne by mosquitoes that have plagued us through history. The historical force of yellow fever, borne by *Ae. aegypti* mosquitoes is, at least, comparable to that of malaria, and the panic it enacted in affected populations likely far exceeded it (Spielman and D’Antonio, 2001). The terror of a yellow fever outbreak is something which we, in the developed world, are thankfully spared. The disease’s symptoms are truly horrific in nature: the illness begins with fever, chills, and muscle pain so extreme that people often believe their bones are breaking. The disease commonly causes blinding headaches and stabbing pains in the eyes. In 15-25% of people it infects, the disease moves
on to a deadly second phase: the liver fails, turning the skin yellow; the eyes turn blood red; blood begins to ooze from the nose and mouth, and the internal hemorrhaging causes blood to flow into the stomach, where it later causes a telltale black vomit. At this stage death is inevitable and follows in the next few days (Monath, 2001). Whilst Max Theiler in 1937 formulated an effective vaccine for yellow fever (Norrby, 2007), there remains no effective treatment today.

Before Walter Reed and Jesse Lazear discovered that yellow fever is vectored by *Ae. aegypti* mosquitoes at the turn of twentieth century, there was essentially no way to combat outbreaks of this disease. Like malaria, up until the twentieth century its origins were a complete mystery. The uncertainty surrounding its transmission amplified the fear it installed in communities where yellow fever outbreaks occurred. Philadelphia in the 1790s was the busiest port in the newly independent United States of America. However, in 1793 the city was paralysed by a disastrous outbreak of yellow fever (see Figure 1.5). Those that survived the first few weeks of the epidemic (including President Washington) fled to the countryside (Spielman and D’Antonio, 2001). The arrival of winter eventually caused an end to the epidemic, by killing the *Ae. aegypti* mosquitoes that vector it, but the human toll was enormous - over 5,000 deaths, roughly 10% of its population at the time, perished (Powell, 1949).

Yellow fever also wreaked havoc with the colonial powers by killing expeditionaries sent to colonise parts of Africa and the New World. Indeed, the legend of the “Flying Dutchman”, a ghost vessel fated to haunt the seas, probably had roots in the fate of real expeditions (Vainio, Cutts, and World Health Organisation, 1998). In a landmark paper, Acemoglu, Johnson, and Robinson, 2000 attribute differences in the economic development of nations today to the differing risks of mortality (mainly due to yellow fever and malaria) faced by European settlers between the 17th and 19th centuries. In countries where the risk
of yellow fever was greatest, the colonial powers created extractive states, typified by the
Belgian Republic of Congo. The purpose of these states was to expropriate the country’s
resources with the minimum investment necessary, and so these states lacked many of the
key requirements for economic development, for example, property rights. In other, less
disease-ridden areas, the Europeans set up “Neo-Europes”, with many of the laws and
institutions that were commonplace in Europe at the time, for example in Australia and
the United States. Acemoglu, Johnson, and Robinson, 2000 argue that these differences
in nascent institutions are still felt today in their daughter societies, and have a real effect
on their rate of economic development.

Whilst there exist other mosquito borne diseases that have affected the course of history
I do not discuss them here, as their influence was likely less extreme than malaria and yellow
fever. However, I argue in Section 1.4, that their importance in the developing world today
is paramount and is a major contributor to the constellation of sicknesses experienced by
those most vulnerable.
Figure 1.3: The geographical distribution of indigenous malaria (black dots) and marshlands (shaded areas) in England from a survey undertaken in 1864. Image reproduced from the Wellcome collection under the creative commons licence and is based on James, 1929.
Figure 1.4: A phylogeny of *Plasmodium spp.* based on mitochondrial DNA. The human malarial parasites are indicated in red. Image reproduced from Pacheco et al., 2013.
Figure 1.5: A scene depicting life and death in Philadelphia during the 1793 outbreak of yellow fever. Image reproduced from Gumm, 2010.
1.4. THE IMPACT OF MOSQUITO BORNE DISEASE IN THE DEVELOPING WORLD TODAY

1.4 The impact of mosquito borne disease in the developing world today

Something is wrong here. War, disease, death, destruction, hunger, filth, poverty, torture, crime, corruption, and the Ice Capades. Something is definitely wrong. This is not good work...I firmly believe, looking at these results, that if there is a God, it has to be a man. No woman could or would ever fuck things up like this.

– The satirist George Carlin, 1937-2008

Before attempting to estimate the costs of malaria to the world today, it is worth remembering that natural selection has already made its own macabre valuation of the disease’s devastating power. J.B.S. Haldane in 1948 was the first to propose the “malaria hypothesis” - that certain genetic polymorphisms, particularly those that affect red blood cells, have been selected for because of the protection they provide against malaria, even though they often confer severe fitness costs to the individual (Haldane, 1949). Haldane originally proposed the hypothesis to explain the high prevalence of thalassemia (an illness causing anaemia) along the Mediterranean coast, where malaria used to be endemic, which, he believed, might make individuals more resistant to malaria infection. An individual that is heterozygous for the thalassemia trait may not experience any symptoms, but homozygotes have a more serious version of the illness, whose treatment involves regular blood transfusions, and historically would have almost certainly died before reaching reproductive age. More recently, Haldane’s conjecture has been confirmed, with evidence demonstrating that heterozygous and homozygous individuals with certain alpha+ thalassemias have a risk of malaria disease that is about one half of normal (Allen et al., 1997). The link between thalassemia and malaria is thought to be so dependable that researchers have used markers on thalassemia genes to trace the migrations of the ancient Greeks, Phoenicians,
and Malayo-polynesians (all civilisations likely plagued by malaria; Cavalli-Sforza, 2001).

There are, however, polymorphisms that impose an even greater risk of mortality than
thalassemia.

The sickle cell trait is associated with a point mutation in the gene for the beta chain
of haemoglobin S, which leads red blood cells to adopt a rigid ‘S’ shape when oxygen levels
in the blood are low (Pauling et al., 1949). The sickle cell trait provides 90% protection
against *P. falciparum* malaria when inherited from a single parent, even though the same
allele, when inherited from both parents, is fatal (Carter and Mendis, 2002). In parts of
Africa where *P. falciparum* malaria is rife, the frequency of the sickle cell allele has been
estimated to be from 15% to 30% (Figure 1.6; Livingstone, 2009; Cavalli-Sforza, Menozzi,
and Piazza, 1994; Piel et al., 2010), whose upper bound would mean that roughly 1 in 10
children are born homozygotes and will likely die soon after birth. Essentially, the risk of
death from malaria is so high that naturally selection allows a regularly fatal mutation to
circulate in the population (Sachs and Malaney, 2002).

The burden of malaria - like for many other infectious diseases - is not distributed
equitably around the globe, driven by the particular vector species present, environmental
factors (Carter and Mendis, 2002), and differences in economic development (Sachs and
Malaney, 2002). Over 90% of malaria deaths occur in Sub-Saharan Africa, the majority
of whom are children under five years old (Murray et al., 2012), and estimates suggest
that there are 0.2 to 0.6 billion cases of malaria occur globally each year (Murray et al.,
2012; World Health Organisation, 2017c). The symptoms of malaria infection take on
many different forms according to which of the five species of *Plasmodium* parasite an
individual is infected with. Common to all infections are periodic episodes of fevers with
chills, sweating, and rigor (Carter and Mendis, 2002). Malaria also has symptoms common
to other infectious illnesses, including headaches, body ache, and nausea. Infection with
1.4. MOSQUITO-BORNE DISEASE IN DEVELOPING WORLD

*P. falciparum* causes the most lethal form of the disease, which can lead to severe anaemia and major dysfunction of the lungs, kidney, liver, and a much-feared brain condition called cerebral malaria. Cerebral malaria is an encephalopathy that causes coma, seizures and life-long neurological problems for those affected (mostly children) and results death in 15% to 50% of patients (Idro, Jenkins, and Newton, 2005; MacPherson et al., 1985).

Overall, malaria causes the premature deaths of between 0.5 and 1.2 million people each year (Murray et al., 2012; World Health Organisation, 2015) - about equal to the combined population of Glasgow and Edinburgh or the total population of Estonia.

Malaria weighs so significantly on affected societies that it traps them in poverty. A cross-country regression analysis that controlled for levels of human capital, life expectancy, initial income, as well as a host of geographical factors, concluded that countries where a high proportion of the population lived in regions of *P. falciparum* transmission had an annual GDP growth rate that was 1.3% lower, on average, than other countries between 1965 and 1990 (Gallup and Sachs, 2001). This implies that two countries - one in a malarial zone, another not affected by the disease - with the same income and population size in 1965, which experienced similar population growth rates over the period, would have a minimum difference in standard of living (measured by GDP per capita) of 38% in 1990. This difference of GDP per capita is comparable to that between Sudan and Vietnam today!

There are also numerous other, indirect, effects of malaria on those societies it plagues. For example, the inflated risk of childhood mortality means that women tend to have more children (Yamada, 1985), meaning that families invest less in each child’s education, and contributes to an already severe gender imbalance in the workplace in the world’s poorest countries (Sachs and Malaney, 2002). Since adults in malaria-affected regions develop partial immunity to the disease, children bear the bulk of the disease burden. Malaria causes the death of many pre-school children, but older children also suffer from its effects,
resulting in missed days of schooling. The effects of this on children’s education are not trivial. For example, a study on absenteeism in Kenyan schools concluded that 11% of school days in primary school are missed because of malaria, and over 4% in secondary school (Leighton and Foster, 1993).

Overall, the global footprint of malaria is enormous and inflicts suffering on those most vulnerable. Unfortunately, there are many other mosquito borne diseases today that also inflict suffering whose importance is growing. The dengue virus is a member of the flavivirus family that includes the viruses that cause yellow fever and Japanese encephalitis (Westaway and Blok, 1997). The pandemic today consists of four viral subtypes that co-circulate. Infection with one of these subtypes confers lifetime immunity to that subtype only and, because of this, individuals in hyper-endemic areas (where all subtypes circulate) often become infected with all four subtypes during their lifetime (Gubler, 1998). Later infection with a second strain that is different to the first is more likely to result in more serious illness (Halstead, 1988).

Dengue virus infection in humans causes a spectrum of symptoms ranging from no noticeable effect to mild fever to severe and possibly fatal hemorrhagic disease (World Health Organisation, 2009). Dengue fever is primarily an illness of older children and adults (Gubler, 1998). The set of symptoms are similar to other febrile conditions (including malaria) that presumably contribute to its misdiagnosis, and includes headaches, nausea, body aches and vomiting (World Health Organisation, 2009). Dengue fever is rarely fatal and typically lasts less than a week (Gubler, 1998), however, dengue infection can cause a much more serious condition known as dengue hemorrhagic fever (DHF). DHF has similar early symptoms as dengue fever but can result in bleeding in the skin, from the gums, and internally, as well as organ impairment (World Health Organisation, 2009). Unfortunately, the serious illnesses caused by dengue infection, including DHF, and another potentially
fatal condition referred to as *dengue shock syndrome*, predominantly prey on children, usually under the age of 15 (Dietz et al., 1996; World Health Organisation, 2009).

The first record of a major outbreak thought to be caused by dengue occurred in the seventeenth century in the Caribbean (McSherry, 1982; Dick et al., 2012), with similar epidemics occurring Asia and Africa probably about a century later (Gubler, 1998). However, accounts of illness similar to that of dengue exist that predate this period. The earliest record found yet is in a Chinese medical encyclopaedia from the Chin Dynasty (265 to 420 AD; Nobuchi, 1979). The disease was referred to as “water poison” by the Chinese and, in an impressive epidemic leap - over one and a half thousand years before Ronald Ross’ discovery that mosquitoes transmit malaria - was thought to connected with flying insects associated with water!

From the emergence of the first likely outbreaks of dengue in the Americas in the 1600s until the 1940s, the pattern of disease were was characterised by relatively infrequent, often large, outbreaks (Gubler, 1998; Dick et al., 2012). The ecological disruption wreaked by the second world war likely created the ideal conditions for the disease to spread. In particular, troop movements probably contributed to expansion of territory of *Ae. aegypti*, which rose in prominence to become, now, the predominant vector of dengue (Gubler, 1998). After the second world war, eradication programmes, fuelled by the recent creation of DDT, resulted in large reductions in the population densities of *Ae. aegypti*, particularly in the Americas (Gubler, 1998), and there was a commensurate fall in cases. The 1970s saw a deterioration of the vector control campaigns, and by the 1990s *Ae. aegypti* had regained much of the territory it held before eradication programmes began (Figure 1.7; Gubler, 1998). Dengue fever inevitably followed on the coat-tails of resurgent *Ae. aegypti* populations, and by the 1980s the Americas were experiencing major epidemics of the disease, in some countries that had been free of the disease for 130 years (Gubler, 1998). Today, the World Health
CHAPTER 1. INTRODUCTION

Organisation estimates that over 40% of the world’s population are at risk of dengue, with 70% of those at risk living in the Asia Pacific countries (World Health Organisation, 2017b). Recent estimates suggest that there may be 100 million symptomatic illnesses caused each year by infection with the dengue virus (Bhatt et al., 2013), resulting in 500,000 cases of DHF (Guha-Sapir and Schimmer, 2005). Dengue fever’s meteoric rise in importance, means that it now causes more death and illness in humans than any other virus carried by arthropods. This group of *arthoviruses* includes West Nile Virus, Japanese encephalitis, and yellow fever, whose impact I discuss now.

It is remarkable that an outbreak of yellow fever in Asia has not yet occurred. The conditions are certainly ripe for the yellow fever virus to emerge, and sustain itself: its predominant vector *Ae. aegypti* is common throughout Asia (Gubler, 2004; Kraemer et al., 2015) and there is evidence that the local population of *Ae. aegypti* can vector yellow fever (Tabachnick et al., 1985). Travellers to many of these countries are not required to provide a yellow fever certificate (Monath, 2001; Jentes et al., 2011), suggesting that the local populace are largely susceptible to the disease. Furthermore, numerous species of monkey are found in Asia, some of which could presumably act as a reservoir for the virus, allowing the disease to become endemic in the region. This has lead to calls for Asian countries to change their vaccination policy for travellers to the region, and heightened vigilance for yellow fever cases (Jentes et al., 2011).

Regardless of the potentially disastrous impact of yellow fever on Asia, the disease already forges a lethal path in South America and Africa, where it may infect 200,000 people annually (Monath, 2001). In regions of Africa where the population is seasonally exposed to the virus, children, who are yet to acquire natural immunity to the disease, are at highest risk of illness (Monath, 2001). Unfortunately, the disease can persist through the dry season because it is vertically transmitted in mosquitoes, meaning a mosquito’s
Table 1.1: The estimated global burden of the major mosquito-borne diseases as of March 2017 (World Health Organisation, 2017a). Note the values in the table represent the central estimates presented by World Health Organisation, 2017a and “NA” indicates data not available.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Annual number of cases</th>
<th>Annual number of deaths</th>
<th>Annual disability-adjusted life years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaria</td>
<td>212,000,000</td>
<td>429,000</td>
<td>NA</td>
</tr>
<tr>
<td>Dengue</td>
<td>96,000,000</td>
<td>9,110</td>
<td>1,892,000</td>
</tr>
<tr>
<td>Lymphatic filariasis</td>
<td>38,464,000</td>
<td>NA</td>
<td>2,075,000</td>
</tr>
<tr>
<td>Chikungunya (Americas)</td>
<td>603,000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Zika virus disease (Americas)</td>
<td>500,000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Yellow fever (Africa)</td>
<td>130,000</td>
<td>500</td>
<td>31,000</td>
</tr>
<tr>
<td>Japanese encephalitis</td>
<td>42,500</td>
<td>9,250</td>
<td>431,552</td>
</tr>
<tr>
<td>West Nile fever</td>
<td>2,588</td>
<td>111</td>
<td>NA</td>
</tr>
</tbody>
</table>

ova, harbouring the virus, has been shown to be able to lie in tree holes (not the primary aquatic habitat of Ae. aegypti) ready to hatch when the rains resume (Monath, 2001; Gubler, 2004). In both South America and Africa, it is the poorest, who work in agriculture in close proximity to the simian hosts of the disease, that are most at risk (Monath, 2001; Gubler, 2004). Indeed, hunters in these regions are known to avoid areas of the forest devoid of the sound of monkeys; this can signal that a recent outbreak of yellow fever has wiped out a local population of monkeys (Spielman and D’Antonio, 2001).

There are many other arboviruses that pose significant threat to swathes of the developing world (see Table 1.1). I do not attempt here to do justice to all of these diseases and, instead, briefly describe two of them: chikungunya and Zika. ‘Chikungunya’ is a word from the language of the Makonde people of southeast Tanzania and northern Mozambique, and translated literally means ‘that which bends up’ (Charrel, 2007) or ‘to walk bent over’ (Pialoux et al., 2007a). The disease is thus called because it is associated with intense pain around the joints, which can incapacitate an affected person (Charrel, 2007; Pialoux et al., 2007a). Chikungunya has been associated with large scale epidemics of illness in the developing world. In 2007, an outbreak of the disease swept through the islands of the Indian Ocean. Reunion has only 770,000 human inhabitants, yet there were 265,000 clinical cases
CHAPTER 1. INTRODUCTION

reported during this epidemic (Charrel, 2007). Like dengue and yellow fever, chikungunya is vectored by mosquitoes of the genus *Aedes*, and in different parts of the world, different species are associated with its transmission (although it is largely vectored by *Ae. aegypti* and *Ae. albopictus*). In the Indian Ocean epidemic, the likely vector species was *A. albopictus* - an aggressive anthropophilic mosquito - whose worldwide prevalence has shifted, and largely, increased over the past 40 years (Charrel, 2007; Kraemer et al., 2015). In an Indian outbreak in 2006, which caused 1.4 million cases of the disease, the main vector was *Ae. aegypti* (Pialoux et al., 2007a), whose increased worldwide distribution has already been discussed.

The Zika virus was first discovered in 1947 in the serum of a rhesus monkey caged in the canopy of the Zika forest in Uganda (Dick, Kitchen, and Haddow, 1952). Whilst a handful of Zika cases were recorded in humans (mostly in Africa), the disease was mostly forgotten about, until an outbreak of illness occurred on a small island in the South Pacific in 2007. In April and May of that year, a number of people on Yap island in the Federal States of Micronesia began to suffer from symptoms encompassing rash, fever and arthritis (Duffy et al., 2009). Attending physicians noted that the disease’s symptoms were distinct from dengue, and serum samples from the patients suggested that the aetiological agent was the Zika virus (Duffy et al., 2009; Hayes, 2009). No further cases of Zika infection were detected in the South Pacific until 2013, when an explosive outbreak occurred in French Polynesia, with an estimated 28,000 patients seeking medical care (Musso, Nilles, and Cao-Lormeau, 2014). The disease then spread through the rest of the South Pacific (Musso, Nilles, and Cao-Lormeau, 2014) and eventually reached Brazil (Faria et al., 2016), where in 2015 an outbreak began in the city of Feira de Santana in the Bahia region of Brazil (Campos, Bandeira, and Sardi, 2015), where recent modelling work has suggested that the virus infected a large proportion of the city’s inhabitants (Lourenco et al., 2017).
By January 2016, at least 30,000 cases of the disease were reported in Brazil (Faria et al., 2016), and tragically the disease is widely believed to behind the recent rise of microcephaly - a birth defect where the baby’s head is smaller than that of a normal child, which can have lifelong neurological impacts - in the region (Mlakar et al., 2016; Oliveira Melo et al., 2016). Zika virus is known to be vectored by mosquitoes of the genus *Aedes* (Hayes, 2009), including *Ae. aegypti* (Campos, Bandeira, and Sardi, 2015; Faria et al., 2016), whose ubiquity in many parts of the world bodes ominously for the future of this epidemic. Further, the Zika virus has been detected in the semen of an infected man (Musso et al., 2015) which supports the earlier suspicion that this disease can be spread sexually (Foy et al., 2011). We are just beginning to understand the magnitude of the Zika epidemic, and the suffering it inflicts on affected individuals. Also, as the disease evolves to avoid our immune responses, it is possible that its virulence will increase (Ewald, 1994) and we could other serious manifestations of Zika infection.

Mosquitoes also carry a range of other diseases, which affect both humans and animals, and this space cannot do these justice (see Table 1.1). One condition which, I believe, has not received adequate attention, is *filariasis*, and I briefly discuss it now. Although this disease does not regularly cause death, it nonetheless inflicts considerable suffering on those unfortunate to become infected with the aetiological filarial nematodes (Taylor, Hoerauf, and Bockarie, 2010; Murray et al., 2013). Filariasis can manifest in *elephantiasis*, which with little doubt, is one of the most stigmatising of tropical diseases. Elephantiasis occurs when the fluid in the lymphatic system becomes blocked, which frequently causes the legs or genitals to swell to enormous proportions (Farrar et al., 2013). As anyone who has ever seen an individual with these symptoms can testify, it is hard to imagine a disease that could cause more gruesome transfiguration in those it affects. Historically, the scientific study of filariasis played a pivotal role in the discovery that mosquitoes could...
vector diseases, and it is this historical context which I describe next.
Figure 1.6: Confirming Haldane’s hypothesis: the distribution of sickle haemoglobin (top) and a historical map of malaria endemicity (bottom). Image reproduced from Piel et al., 2010.
Figure 1.7: The distribution of *Ae. aegypti* mosquitoes in South America before (left), immediately after (middle), and long after (right) the campaign to eradicate this species from the region. Image reproduced from Gubler, 1998.
1.5 Mosquitoes as vectors

The etymological origin of the word “malaria” is rooted in nineteenth century Italy. At the time, it was widely believed that fever was caused by a “miasma”, or poisoning of the air. Indeed, the Italian “mal’aria” literally translated means “bad air”, and this became a popular term for the disease in this period (Snowden, 2008). The path by which mosquitoes came to be implicated in the transmission of malaria is hotly contested, and is a source of national pride for some historians of science. Here, I outline a brief narrative of history described by Spielman and D’Antonio, 2001, although encourage interested readers to consult Ross, 1923 and Grassi, 1924 for two different perspectives on this subject.

The discovery that malaria is transmitted by anopheline mosquitoes actually begins with research into lymphatic filariasis. Elephantiasis made a strong impression on Dr Patrick Manson, a British medical officer stationed in Taiwan (then called Formosa, which was part of China) in the 1870s. In 1875, on a holiday from his service in China, Manson visited London and bought a compound microscope. When he returned to China, he set about investigating the causes of elephantiasis, at first, by conducting autopsies in secret (at the time this practice was outlawed in China) of bodies of patients who had died whilst suffering from this disease. In one patient, Manson found threadlike worms, about the length of a person’s arm, coiled inside within their lymph nodes. In an experiment of questionable ethics, Manson recruited his gardener, Hinlo, whose blood teemed with the filarial parasites, to try to demonstrate that the pathogen could be transmitted from man to mosquito. Hinlo slept within a tent Manson dubbed the “sleeping house” (how much sleep Hinlo got that night is not recorded) whilst *Culex quinquefasciatus* mosquitoes (also known as the southern house mosquito) fed on him. During the next five days, Manson dissected the recaptured mosquitoes and was able to find larval stages of the filarial parasite in some of the specimens. Manson’s achievement was monumental - he was the first to
experimentally demonstrate the role of mosquitoes in spreading infectious disease!

Around the same time that Manson was performing his filariasis experiments, a French Army doctor named Charles Laveran was stationed in a malaria ridden outpost of the French Foreign Legion in Algeria. Out of concern for the military, not the colonial subjects, the French government had sent Laveran to Algeria. Here Laveran was free to perform autopsies and undertake his own experiments. In November of 1880, Laveran drew blood from a patient suffering from extreme fever and placed a droplet of the blood on a microscope slide. Under the magnification, he could see that the blood was alive with tiny protozoan creatures, which he called *Oscillaria malariae* (Cox, 2010). These were the first observations of the malarial parasite, which would later be renamed as *Plasmodium*, and resulted in Laveran being awarded the Nobel Prize in 1907.

Laveran’s discovery is rightly lauded for its elucidation of the causes of malaria. What remained unclear at the time, however, was how the disease was transmitted to humans. This required the work of another colonial doctor, Ronald Ross, who was serving in India during the 1880s and 1890s. Ross was born in the Himalayas as the eldest of a British general’s ten children. He was a true polymath and enjoyed a passion for mathematics and poetry as a child. At the age of twenty-seven, having recently been admitted as a member of the Royal College of Surgeons, Ross was posted to Bangalore.

In 1894 in London, Ross was on a break from his service in India and approached Patrick Manson, who was then famous for his work on filariasis. Ross shared with Manson his theory (others, decades before Ross, had also had the same theory) that malaria was also vectored by mosquitoes and Manson agreed with him. Spurred on by Manson, Ross boarded a boat back to India. Ross was so excited about the prospect of finding malarial parasites in humans that he apparently examined the blood of all passengers on the vessel. An excitable graduate student can certainly appreciate his enthusiasm.
Interrupted by major cholera outbreaks on the subcontinent, Ross’ work continued, without success, for three years. During this time, Ross carried out numerous dissections of a grey mosquito - probably *Culex pipiens* (the common house mosquito) - and a brindled one - probably *Ae. aegypti*, the carrier of yellow fever, dengue and Zika - and in all cases failed to find malarial parasites. It was not until 1897, when his lab assistant raised a number of larvae that matured into “big brown fellows” - a mosquito species which Ross had not seen before (anophelines) - that Ross would find success. Ross fed this mosquito species on malarious patients then set about dissecting the mosquito specimens a few days after the initial infection. On examining the gut of one, Ross saw dark, pigmented cells that were clearly not part of the mosquito itself. These observations were validated by similar dissections the following day. Ross had done it - he had found traces of the malarial parasite in mosquitoes for the first time! In 1897, work by the Italian scientists Bignami, Golgi (who won the Nobel prize for his work on the subcellular compartment which is now known as the “Golgi apparatus”), and Grassi amongst others closed the transmission cycle. They did this by infecting mosquitoes with malaria by feeding them on febrile human subjects, and then produced an infection in uninfected humans, by feeding the infected mosquitoes on them.

Following hotly on the success of Ross and others in demonstrating the anopheline mosquito’s role in transmitting malaria, the mystery of the epidemiology of yellow fever was elucidated, again by the work done by an army physician. In 1900, Walter Reed was stationed on Cuba with his assistant Jesse Lazear, following the occupation of the island by their native US at the end of the Spanish-American war. In a series of experiments that would claim the life of Lazear at age 31 (after he purposely was bitten by a mosquito which had fed on the blood of yellow fever patients), the work of the pair demonstrated that *Ae. aegypti* was able to transmit the disease. The definitive experiment, performed a month
after Lazear’s death, involved “volunteers” (this experiment preceded the discovery of the yellow fever vaccine by 37 years) sleeping in one of two tents: the less lucky volunteers slept in a tent with mosquitoes which had fed on patients suffering from yellow fever; the slightly more fortunate individuals slept in another tent, with blankets soiled with the telltale black vomit of yellow fever victims. Only those in the mosquito tent became ill, confirming that mosquitoes were typically required for transmission. Whilst a Pyrrhic victory for Lazear (and the unlucky participants of the experiment), this discovery was crucial in shaping vector control efforts that would soon be effective in curbing the incidence of this terrible disease. Despite his crucial part in uncovering the mosquito’s role in yellow fever, Reed did not live much longer than many of his volunteers, and died in 1902, when his appendix ruptured.

1.6 Disease control since Ronald Ross

The head of that statue was made of fine gold, its breast and its arms of silver, its belly and its thighs of bronze, its legs of iron, its feet partly of iron and partly of clay. You continued looking until a stone was cut out without hands, and it struck the statue on its feet of iron and clay and crushed them...Then the iron, the clay, the bronze, the silver and the gold were crushed all at the same time and became like chaff from the summer threshing floors; and the wind carried them away so that not a trace of them was found.


Following the discovery of the mosquito’s role in malarial transmission, leading Malarialogist Giovanni Grassi declared malaria was a “giant with a feet of clay”, a biblical
Figure 1.8: The occurrence of phrases associated with vector borne diseases (the black dashed line represents DDT) in books published in the English language. The corpus of books from which these quantities are derived is the “English 2012” Google nGram corpus, compiled from digitised books by Google. The requisite data were collected using the ngramr R package (Schmidt and Heckendorf, 2017), and represent two period moving average percentages of all n-grams (words for single words; pairs of words for two words) for all books published in that year.

reference (from the book of Daniel, whose original quote appears at the beginning of this section), meaning an obstacle which can be readily overcome (Sachs and Malaney, 2002).

Sadly, more than a century on from Grassi’s declaration, malaria still stands firmly (albeit less firmly than 15 years ago (Bhatt et al., 2015) and much less firmly than 100 years ago (Snow et al., 2017)), its feet presumably made of sterner stuff than soil. Over the
course of the twentieth century, there have been a few unbridled successes, many partial
successes and, unfortunately, a fair number of failed attempts to curtail this disease. What
is often lost in historical descriptions of “failed” attempts to control malaria is that, whilst
the ultimate goal of globally eradicating the disease may have not been met, these con-
siderable efforts were not in vain; many people are now either alive, or live comparatively
better off lives, as a result of them. Much of this section considers the global efforts to
control malaria but, due to these vector borne diseases’ joint reliance on the mosquito and,
importantly, the measures used to control them, their history is, to some degree, shared,
which is illustrated in Figure 1.8. This figure shows the annual counts of a few terms
relating to vector borne diseases, including malaria, from 1880 to 2008, as a percentage
of all words published that year, derived from a corpus of books published in the English
language. Whilst this diagram should not be taken too literally, as it is certainly the case
that word usage does not directly translate into disease incidence, it nonetheless provides
an interesting view of the history of vector borne disease over the period. In this figure,
it can be seen that, whilst malaria and yellow fever (the red and turquoise lines) in 1880
occupied similar space in the printed pages of the time, in the twenty-first century malaria
is more popularly discussed. Also, this figure displays the meteoric rise in discussion of
mosquitoes (the gold line) after Ronald Ross’ work in 1897 demonstrated that malaria can
be transmitted by mosquitoes. Leading up to each of the world wars (indicated by the
blue shaded rectangles), there was increased discussion of the selected diseases, possibly
as English-speaking countries discussed ways to control the disease to give their troops
advantage. Finally, the black dot-dashed line illustrates how the insecticide DDT became
popular following its invention in the 1940s, after which it became the centrepiece of cam-
paigns to control yellow fever in the Americas (Packard, 2007) and a global programme to
eradicate malaria (Packard, 2007; Spielman and D’Antonio, 2001), which persisted until
the 1970s. As discussed, it is dangerous to ascribe movements in each of these series to
the burden of disease or the adoption of certain interventions, however, I cannot resist
making one observation: after the introduction of DDT (the black dot-dashed line), there
appears to have been a substantial reduction in all the mosquito borne diseases (the red,
green, turquoise and mauve lines). This somewhat derivative observation is supported by
considerable evidence (see Figure 1.9 and Packard, 2007 for a summary of this evidence),
and so I believe that this slight misuse of the data are warranted. Though DDT use is
extremely controversial (the second peak in this series around 1970 to 1975 likely reflects
discussion of its possible negative impact on humans and ecosystems; see my discussion be-
low) and its use is banned in many countries, 380 scientists, including three Nobel laureates
in medicine, across 57 countries signed a letter endorsing continued use of DDT (Taverne,
1999; Roberts, Manguin, and Mouchet, 2000).

Not wishing to venture into the politics of DDT usage here, I instead attempt to provide
a history of mosquito borne disease control in the twentieth century, which, I believe is
more or less supported by the data from Figure 1.8. Whilst the discovery of Ross concerned
malaria, the first major victory against over mosquito borne disease was actually claimed
against yellow fever, which I now discuss.

After Reed and Lazear demonstrated that \textit{Ae. aegypti} mosquitoes were responsible for
transmitting yellow fever, another US army man, also stationed in Cuba, took up the task of
eradicating the disease from the island’s capital, Havana (Spielman and D’Antonio, 2001).
Major William Gorgas, an uncompromising man whose temperament would later cause
him trouble in Panama, organised brigades of soldiers to patrol the streets under orders to
search and destroy containers containing standing water, to which the aquatic stages of \textit{Ae.
aegypti} are adapted. Ponds were also oiled each week and residents punished if mosquito
larvae were found on their premises. One of the first vector control campaigns in history
Figure 1.9: The changing distribution of malaria from 1946 to 1994. The distribution of disease in 1946 is before the global push for eradication began, whereas the map for 1966 shows prevalence in the midst of the campaign, and for 1994, the same, although long after it ended. Image reproduced from (Sachs and Malaney, 2002).

(see World Health Organisation, 2017a Annexe 2 for a list of other such campaigns) was a great success; after only five months, yellow fever was eradicated from Havana (Spielman and D’Antonio, 2001).

After his victory over yellow fever in Cuba, Gorgas was deployed to Panama, where America was in midst of attempting to complete the Panama Canal. In their efforts to construct the Panama canal, the French had lost some 20,000 workers, mainly to yellow fever and malaria (Packard, 2007). The Americans, who had taken over work on the canal in 1902, at first fared no better. In 1904, when Gorgas arrived in the country, he began a militaristic campaign, of a nature similar to that he had used in Cuba. Specifically, Gorgas divided the canal zone into 25 districts, each of which was placed under the control of
a sanitary inspector who oversaw between 20 and 100 labourers (Gorgas, 1915). Having
identified the breeding habits of local malaria and yellow fever carrying mosquitoes, the
teams set about destroying their aquatic habitats. The teams cleared the French gardens
and water receptacles away from living quarters, and spread lava killing oil on larval aquatic
habitats (Spielman and D’Antonio, 2001). He also sent soldiers into homes and barracks to
swat adult mosquitoes, one by one (Spielman and D’Antonio, 2001). To kill anophelines,
the teams drained marshy areas and filled them in with cement (Packard, 2007). Gorgas’
uncompromising nature won him few friends in the commission overseeing the construction
of the canal, who even tried to replace him at one point (Spielman and D’Antonio, 2001).
Fortunately for Gorgas, his success in Cuba had won him friends in high places, and the
then President of the United States, Theodore Roosevelt, decreed that Gorgas should
continue his work in Panama (Spielman and D’Antonio, 2001). (It is difficult to conceiv
of a field entomologist currying such favour today.) By 1906, Gorgas had managed to
eradicate yellow fever from the canal zone (Spielman and D’Antonio, 2001), and malaria saw
significant declines over the ensuing few years. Malaria was presumably harder to suppress
because of the greater difficulty of targeting the anopheline mosquito vectors (Spielman
and D’Antonio, 2001). Nonetheless, deaths from malaria declined from a maximum of
about 16 per 1000 workers in July 1906, to a minimum of less than 3 per 1000 in December
1909 (Packard, 2007).

During the 1910s and 1920s, malaria control was split between various approaches,
including vector control (primarily using the larvicide Paris Green; Packard, 2007), tar-
geting the parasite through quinine and social and economic development programmes
that emphasised the human ecological aspect of disease transmission. Schemes aimed at
social and economic betterment were pioneered in Italy after the First World War, where
so-called bonifica integrale or bonification policies were instituted. The most notable suc-
cesses of bonification were achieved under Mussolini’s facist government, who managed to drain the Pontine Marshes and eradicate malaria from a province previously gripped by malaria (Spielman and D’Antonio, 2001; Packard, 2007). The scheme began in 1929 and resulted in considerable areas of land being reclaimed for agriculture. This stimulated a large scale immigration to the area, which saw its population rise from 1637 people in 1928 to 60,000 in 1939 (Packard, 2007), largely due to those seeking employment in the newly available agricultural land. Along with land reclamation efforts, vector control (principally by attacking the larval stages of the mosquito using Paris Green) and quinine were integral parts of the project. Bonification schemes also focussed on the societal elements of disease transmission, and included installing high quality housing and setting up regional health clubs. Interestingly, these human aspects of disease control - notably the construction of mosquito-proof houses - have recently proven to be a growing research areas (Lindsay, Emerson, and Charlwood, 2002), with a meta-analysis estimating that individuals living in modern housing had 47% lower odds of contracting malaria compared to traditional housing (although the authors recognised that there was likely sampling bias in the individual studies; Tusting et al., 2015).

The emphasis of human and drug-based approaches to malaria was reflected in the discourse of the time. The League of Nations, founded after the end of the First World War, had a department called the Malaria Commission, whose role was to deal with the increased disease prevalence of this disease during and following the war (Packard, 2007). The first report in 1924 by the British Malariologist, Colonel Sydney James, emphasised the importance in “quininization”, housing, education, nutrition and health in disease control. It has been suggested that its aversion to vector control may have stemmed from James’ own personal experience in India, where Britain failed to eradicate malaria at Mian Mir, near Lahore (which are both in present day Pakistan; Packard, 2007; Bynum, 1994).
Whilst in the early 1930s, the prevailing attitude swung away from vector control, this was about to change, spurred on by the notable success of Fred Soper in Brazil and, later, the discovery of the potent insecticide *dichloro diphenyl trichloroethane*, better known by its acronym, *DDT*.

In 1930, Raymond Shannon, an entomologist from the Rockefeller Foundation, was taking a walk from his home in Natal, Brazil, when he noticed strange larvae in a puddle. Shannon immediately recognised the danger of this small organism - he had found the major African malarial vector *Anopheles gambiae* on Brazilian soil - and alerted the Brazilian authorities. Apart from a few minor outbreaks in the intervening years, the impact of *Anopheles gambiae* on the country was relatively minor until 1938, when a major outbreak of malaria struck in the north-east of Brazil, resulting in over 100,000 disease cases (Spielman and D’Antonio, 2001). Whilst *A. gambiae* no doubt played an important role in this outbreak, it was also fuelled by a swollen population of migrant farmers from the interior, where a devastating drought was raging (Packard, 2007). Reacting to this public health disaster, Brazil’s president created an antimalarial response unit staffed with 4000 workers, and placed it under sole control of a Rockefeller associated named Fred Soper. Working with the authority of a war General, Soper launched a war on *A. gambiae* in Brazil, and exercised military rigour in his campaign. Cars were fumigated whenever drivers arrived at checkpoints and ships were boarded and investigated for *Anopheles* infestation (Spielman and D’Antonio, 2001). The most powerful weapon in Soper’s arsenal was the larvicide *copper acetarsenite*, better known as *Paris Green* because of its vivid colour which, at the time, was associated with the Impressionist art of the French Riviera. Soper directed the extensive use of Paris Green, which exacted a heavily toll on *A. gambiae* populations. At its peak, this African mosquito held terrain in over 18,000 square miles of territory over three Brazilian states (see Figure 1.10; Spielman and D’Antonio, 2001), yet in a single year,
Soper had virtually eliminate the vector from the country!

Figure 1.10: A map of the distribution of *Anopheles gambiae* and physical geography in Brazil c.1932. Image reproduced from (Soper and Wilson, 1943).

After Fred Soper’s Brazilian victory over mosquito, vector control was now back at the centre of efforts to control mosquito borne disease before the start of the Second World War, and a wartime discovery was about to cement its position further. In the summer of 1943, a colleague of Soper’s returned from Algeria with news of a new pesticide - then referred to as Neocide - which had demonstrated great efficacy in vector control (Spielman and D’Antonio, 2001). Soper subsequently travelled to Algeria and conducted trials on Axis
prisoners, who were successfully deloused with the chemical (Kinkela, 2011). This chemical, hitherto known by its chemical acronym DDT, was to become an integral part of the Allied war effort, with British and American factories manufacturing tonnes of it throughout the war period. Ironically, the chemical was actually derived from a German moth killing pesticide, which originated over fifty years beforehand (Spielman and D’Antonio, 2001).

At the end of the war, a field trial of this promising pesticide DDT was conducted on the island of Sardinia, orchestrated by a partnership of the Italian government and the Rockefeller Foundation. Throughout the campaign, an extensive workforce sprayed over 2 million rooms in over 300,000 houses with DDT, and also vast quantities of the chemical were dropped on fields by Italian Air Force helicopters and planes (Spielman and D’Antonio, 2001). The campaign was a success - malaria was eradicated from the island - but, it was not unqualified because, after five years of effort and several million dollars spent, the malaria vector Anopheles labranchiae could still be found, which was, unfortunately, a foretelling of what was to follow.

In 1955, member states of the recently established World Health Organisation convened in Mexico City to vote on an ambitious proposal: a plan to globally eradicate malaria (known as the Global Malaria Eradication Programme (GMEP)). The linchpin of the proposed strategy lay in three letters: DDT. The proposal was passed, and over the next 14 years, over $1.4 billion was spent on the eradication programme, the majority of which was spent on DDT (Spielman and D’Antonio, 2001). So strong was the faith in DDT that following the announcement of the eradication plan, grants for research in malariology dried up virtually overnight, and some universities, Harvard, for example, stopped teaching the subject entirely (Spielman and D’Antonio, 2001). When the eradication programme was passed, most believed that the elimination of disease in sub-Saharan Africa, where the bulk burden of disease lay, would follow after a sustained effort (Packard, 2007). Unfortunately,
CHAPTER 1. INTRODUCTION

this proved to be wildly optimistic - the campaign was successful in eliminating malaria from mainland Europe and the southern United States, but in those worst affected parts of the world, the disease stubbornly remained (see Figure 1.9). Before I discuss the reasons behind the failed attempt to eradicate malaria from the face of the globe, I think it is important to stress the campaign’s successes. Hundreds of thousands, possibly millions, of people have lived longer, less wretched lives as a direct result of the malaria eradication programme and, today, peoples across many countries likely enjoy a higher living standard because of it. The fact that the campaign did not live up to its ambitious expectations is disappointing, but an argument can nonetheless be made that this programme was one of the most impactful public health campaigns of the twentieth century.

Why didn’t GMEP meet its goals? There were a complex nexus of factors that contributed to this outcome, including waning international support following negative publicity for the use of DDT (after Rachel Carson published her polemic, *Silent Spring*, against the use of the insecticide in 1962), the failure to recognise local ecological and epidemiological variability (Nájera, 2001), the lack of research in mosquito vectors and malariology in general following the campaign’s initiation (Nájera, 2001; Nájera, González-Silva, and Alonso, 2011), the loss of the US as a crucial programme donor in 1963 (Nájera, 2001), climatic events (Packard, 2007) and resistance to chloroquine - a major antimalarial drug which is still used today. Also, by design, the programme did not realistically aim to achieve global eradication since it largely ignored Sub-Saharan Africa - a region contributing the majority of the world’s disease cases (Nájera, 2001). Indeed, when the programme was approved at the World Health Assembly in 1955 in Mexico, Sub-Saharan Africa was represented by only three independent countries: Ethiopia, Liberia and South Africa, and a handful of metropolitan governments of colonial powers (Nájera, 2001). Rather than discuss in detail all of the causes of GMEP’s failure to achieve global eradication (see Nájera,
DISEASE CONTROL SINCE RONALD ROSS

I focus on another leading cause of the programme’s dispiriting results - mosquito resistance to DDT. The campaign relied heavily on the indoor spraying of the walls of houses with DDT, based on the assumption that female anophelines predominantly feed indoors and, afterwards, rest on the indoor walls. Unfortunately, whilst, on average, these claims were true, there were considerable exceptions to them, meaning that if the endophilic mosquitoes died because of the chemical, they were likely replaced by other, more exophilic, anophelines (Packard, 2007). Unfortunately, mosquitoes also evolved biochemical resistance to the pesticide, meaning that resistant vectors still fed on humans whilst they slept and rested on the indoor walls afterwards. Because of the experiences in Sardinia, where some mosquitoes remained following the onslaught of DDT, as well as similar experiences in Greece and Saudi Arabia, DDT resistance was actually known about before the start of the campaign and, because of this risk, “TIME IS OF THE ESSENCE” was written in capital letters on the original plan (Spielman and D’Antonio, 2001). Unfortunately, by 1969, when the WHO ostensibly pulled the plug on the eradication programme (and refocussed it towards disease control), this time had run out. Whilst the campaign ended, its legacy likely persists to this day, in some parts of the world where endemic malaria is now lost from memory.

In the 1980s and 1990s, malaria control strategy veered away from vector control and focussed on approaches that target the human stages of the parasite (Godfray, 2013), for example, using the highly effective drug artemisinin (recently discovered by Western science, although it was in use in Chinese medicine for over 2000 years before this; Klayman, 1985; Hsu, 2006) and efforts to develop a vaccine. Over the last twenty or so years, there has been a shift back towards vector control, perhaps driven by the difficulty in developing vaccines against the complex life stages of the malarial parasite (Godfray, 2013), and the growing presence of other mosquito borne diseases, like dengue fever (Gubler,

2001; Nájera, González-Silva, and Alonso, 2011 for a thorough analysis of GMEP),
CHAPTER 1. INTRODUCTION

Whilst these vector control efforts (catalysed by the establishment of the Millennium Development Goals in 2000) caused unprecedented declines in disease cases from 2000-2015 (Bhatt et al., 2015), it seems less certain that this trend will continue. More recently, the downward trend in case numbers appears to have flattened; in 2016, it has been estimated that there were 5 million more disease cases than in 2015 (World Health Organisation, 2017c). The reasons behind this growth in cases remain unclear, although it is feasible that donor fatigue, intervention coverage gaps (for example, fewer insecticide treated nets were delivered in 2016 than in 2015; World Health Organisation, 2017c), climate change, insecticide resistance and parasitic drug resistance could have contributed. Indeed, it seems there is still much we do not know about vector borne disease and its control and, importantly, the mosquito itself.

1.7 Why we need to understand mosquito vectors

...nature is almost infinite in her variety, which means that we can make very few absolute declarations about mosquitoes.

– Mosquito, 2002, Andrew Spielman and Michael D’Antonio

A key characteristic of any infectious disease is its basic reproduction number, $R_0$, which measures the ability of a disease to propagate in a population. Specifically, $R_0$ is the number of new cases generated, on average, from a single case, throughout the period of infection. If $R_0 < 1$, then, on average, the disease does not generate enough subsequent cases to replace those individuals that either recover or die from the infection, and so the prevalence of the disease declines over time. By contrast, if $R_0 > 1$, then the disease will spread through the population. Where $R_0 = 1$, the prevalence should remain constant.

As I have already discussed, mosquitoes act as vectors for a number of diseases that
1.7. WHY WE NEED TO UNDERSTAND MOSQUITO VECTORS

affect humans. Whilst I focus on malaria here, it is important to note that many of the epidemiological elements for malaria are common to all of these diseases: an uninfected mosquito can become infected itself by biting an infected host - be they human or animal, although in some cases it may receive the disease from its parents (yellow fever, for example, can be vertically transmitted); in all human-affecting mosquito borne diseases, a mosquito transmits the disease by biting an uninfected host, when the pathogen passes from the mosquito into the human. The centrality of the mosquito to both of these processes - its initial infection and the transmission of infection to a subsequent host - underscores why knowledge of the mosquito’s ecology and behaviour are so crucial for understanding the epidemiology of these diseases.

The epidemiological details of malaria determined by Ronald Ross were put on a mathematical footing by Macdonald, 1957, who devised a mathematical model of the disease. In doing so, the authors made a number of assumptions, which have come to be known as the “classic assumptions” (Smith and McKenzie, 2004), which are useful idealisations that allow development of a simple model, which nonetheless captures many of the key aspects of the epidemiology of malaria. In particular, it is assumed that the mosquito population is homogeneous and of constant size, that mosquitoes do not experience senescence, that is, age-related death, that mosquitoes bite humans uniformly and at random, and that an infected mosquito never becomes uninfected. The model of Macdonald, 1957 results in an expression for $R_0$ given by the following,

$$R_0 = \frac{ma^2bce^{-gn}}{rg},$$

where $m$ is the density of mosquitoes per human, $a$ is called the “human feeding rate” and represents the expected number of bites a mosquito will make on humans per day, $b$ is the probability that an uninfected human becomes infected after being bitten by an infectious
mosquito, $c$ is the probability that an uninfected mosquito becomes infected as the result of a bite on an infected human, $g$ is the “force of mortality” and represents the per capita death rate of a mosquito, $n$ is the duration of the extrinsic incubation period (EIP) - the period of time it takes the ingested \textit{Plasmodium} parasite to transmute and migrate from the mosquito’s gut to its salivary glands, where it can be passed onto another host when the mosquito bites them - and $r$ is the rate of recovery of an infected human from the disease. In what follows, I discuss the sensitivity of malaria transmission to each of the mosquito-related elements of expression (1.1).

It may appear surprising that $R_0$ is least sensitive to the number or density of mosquitoes in a region, since $m$ enters linearly into expression (1.1). This does not mean that this aspect is unimportant for malaria transmission. In many parts of the world, there is significant seasonality in malaria cases, driven by changes in the mosquito vector density. However, expression (1.1) does indicate that substantial changes in the density of mosquitoes are required to significantly affect transmission, and suggests that vector control methods that also affect other parts of disease transmission may be advantageous.

Malaria is vectored by mosquitoes of the genus \textit{Anopheles}, and in Sub-Saharan Africa, where the majority of deaths occur, the principal vector is the \textit{A. gambiae} mosquito. The endemicity of malaria in large parts of Sub-Saharan Africa is likely due mainly to the behavioural characteristics of this species of mosquito (Carter and Mendis, 2002): it favours biting of humans, and the highly domestic habits of \textit{A. gambiae} mean that there is a great deal of contact between them and humans (Coluzzi, 1999). Outside Africa, the anopheline mosquito vectors are zoophilic rather than anthropophilic (Bruce-Chwatt, Garrett-Jones, and Weitz, 1966), meaning that they prefer to feed on animals rather than humans. These vectors of malaria typically have an anthropophilic index (the proportion of blood meals taken on humans) of around 50%, but often much lower, between 10% to 20%
1.7. \textbf{WHY WE NEED TO UNDERSTAND MOSQUITO VECTORS} \hfill 67

(Bruce-Chwatt, Garrett-Jones, and Weitz, 1966). By contrast, \textit{A. gambiae} and \textit{A. funestus} - the two predominant vectors of malaria on the African continent - have anthropophilic indices around 80\%-100\% and 60\%-100\%, respectively (Bruce-Chwatt, Garrett-Jones, and Weitz, 1966). The prejudices of a mosquito’s specific taste for blood matter, because for malaria to be transmitted, a mosquito must first bite an infected human, then later bite another uninfected person. Because two bites are required for disease transmission to occur, transmission has a strong sensitivity to the specificity of mosquitoes towards biting humans. In particular, the effect size depends on the square of the proportion of meals taken on humans - $a^2$ in expression (1.1). Consider a mosquito vector with an anthropophilic index of 10\%, for example, \textit{A. sinensis} (a vector of malaria in the far east; Bruce-Chwatt, Garrett-Jones, and Weitz, 1966). For \textit{A. sinensis} to generate the same level of disease transmission as \textit{A. gambiae}, its population density would need to be 100 times as great!

Why do some mosquito species in Africa (particularly, certain members of the \textit{A. gambiae s.l.} complex) tend to prefer an all-you-can-eat human blood buffet, whereas those outside prefer to feed mainly on the blood of animals? It is possible that this is due to differences in the animals that were available for domestication between Africa and the rest of the world. In Sub-Saharan Africa, there were no animal species that could be readily domesticated, whereas in Europe, Asia, and the Middle East, many animals, including cows, pigs, sheep and goats were domesticated during the Neolithic agrarian revolution around 4,000 to 5,000 years ago (Diamond, 1998). This meant that, whilst in Europe, population densities of animals of a few select species grew enormously, in Sub-Saharan Africa, the animals were more diverse and at lower population densities. This may be why certain anopheline species in Sub-Saharan Africa adapted to become such efficient human blood feeders, since we were the animal species that was available in the highest population densities. In other parts of the world, the mosquito species may have adapted to feed on the
domesticated animals which were available in sufficient numbers and, perhaps, easier prey (Carter and Mendis, 2002). An alternative explanation is that, since humans evolved out of Africa, they have been exposed to the mosquito species there for longer than anywhere else, allowing certain members of the \textit{A. gambiae} complex more opportunity to evolve to exploit humans.

Whilst the $a^2$ term appears simple, it hides subtle complexities, which are crucial for developing effective interventions against malaria. Because of the sensitivity of $R_0$ to the rate of human biting, many vector control methods explicitly target reductions in this aspect of disease transmission. All anopheline vector species feed predominantly indoors at night (Gatton et al., 2013), which partly led to the widespread use of indoor residual spraying of insecticides (IRS), as well as the massive increase in the use of long lasting insecticidal nets (LLINs), since the turn of the twenty-first century. Both of these interventions have been credited with causing the first sustained decline in malaria incidence in a generation (Bhatt et al., 2015), however, a consequence of spraying insecticides indoors is that this gives a selective advantage to those mosquitoes that feed on humans earlier in the day, when they are outdoors. Indeed, exophily (the preference of mosquitoes to rest outdoors after they have fed) is believed to be a key reason why indoor residual spraying failed to control malaria in the Garki project in northern Nigeria (Gatton et al., 2013; Molineaux, Gramiccia, and World Health Organisation, 1980). The sensitivity of $R_0$ to the human biting rate also highlights the importance of understanding the community ecology of vector species (Godfray, 2013). In parts of Africa, the widespread use of LLINs, which provide a barrier between a sleeping person and mosquito, has lead to a shift in vector dominance from the endophilic \textit{A. gambiae sensu stricto} to the more exophilic \textit{A. arabiensis} (Gatton et al., 2013). To better plan effective interventions, it is essential to know the local distribution of species present, and recognise if it is likely that vector
substitution may occur as a result of targeting one of these species.

Disease transmission, as measured by $R_0$, is most sensitive to mosquito mortality, whose effect is quantified by the $e^{-gn}$ term in expression (1.1). Why is disease transmission so sensitive to this aspect of mosquito ecology? Because a mosquito that lives longer, has more opportunity to bite an infected human and become infected itself, then to survive the extrinsic incubation period ($n$; EIP) to allow it to become infectious, and finally has enough time to find and bite an uninfected host. Also, more indirectly, mosquito lifespan influences disease transmission because mosquitoes that live longer produce more offspring and, hence, support larger populations of mosquitoes. The sensitivity of $R_0$ to mosquito mortality means that it is crucial to determine how long they live in the wild.

The predominant approaches to estimating how long mosquitoes live in the wild are MRR experiments and female mosquito dissection-based approaches. In MRRs, lab-reared or wild-caught mosquitoes are marked, typically with a fluorescent dye, and released into the wild. Over days subsequent to the release, collections are made of mosquitoes and the number of marked mosquitoes recaptured is recorded. In populations of mosquitoes with shorter lifespans, the number recaptured will tail off more quickly, because fewer marked mosquitoes persist in the population. Typically, these experiments do not record information on where the mosquitoes were recaptured, and so the rate at which mosquitoes disperse out of the study area also affects the number of mosquitoes available for recapture. This means that estimates of lifespan determined by this method are more accurately interpreted as “time spent in the study area”. In Chapter 2, I provide a more in depth appraisal of this approach to estimating mosquito lifespan and so do not discuss this further now. In dissection-based studies, wild-caught female mosquitoes are dissected in an attempt to estimate their age. This is possible because it is thought that female mosquitoes undergo predictable changes in their reproductive anatomy throughout their lifespan, which can be
used to estimate the age of a given specimen (Polovodova, 1949; Detinova, 1962). In Chapter 3, I describe this approach in more detail and so do not discuss it further now.

The most predominant and impactful interventions - LLINs and indoor residual spraying (IRS) - are thought to be so successful because they each simultaneously target multiple elements of the disease transmission process. Most importantly, it is hoped that the toxic effects of these chemicals kill mosquitoes and, hence, shorten the mean lifespans of populations. By shortening the lifespan of mosquitoes, this likely leads to a decline in the density of mosquitoes that are supported in a given location, again reducing disease transmission. Finally, LLINs (and probably, to a lesser degree, IRS) reduce the effective contact time between mosquitoes and humans, leading to lower biting rates.

Whilst there is widespread consensus that LLINs and IRS have been effective means of vector control, there still exists considerable uncertainty in their mode of action (Gatton et al., 2013). The insecticides used in LLINs and IRS are thought to have threefold effects on vector populations: toxic chemical action, spatial deterancy, and contact irritancy (Smith and Webley, 1969; Lines, Myamba, and Curtis, 1987; Takken, 2002). Whilst there is considerable evidence in support of each of these factors, there is now widespread prevalence of mosquito vector resistance to the four main classes of insecticide which are widely used (Ranson et al., 2011; Ranson and Lissenden, 2016), which threatens to weaken the potency of these methods of vector control. Indeed, results from experimental hut trials demonstrate that mosquito mortality is considerably lower in areas with higher levels of insecticide resistance (N’Guessan et al., 2007). Aside from physiological resistance to the chemicals, there is reasonable evidence that there has been behavioural adaptation to the widespread use of bednets and insecticide spraying (see Gatton et al., 2013 for a summary of the current evidence). This growing resistance means that understanding the mechanism by which certain interventions are effective is essential to be able to design new interventions
An element of expression (1.1) that I have not yet discussed is the EIP, to which disease transmission is almost as sensitive as mosquito lifespan. Rather than being a constant, the EIP for malaria has long been known to be temperature sensitive (Boyd, 1949; Macdonald, 1957; Detinova, 1962), and is reasonably well approximated by the downward-sloping “Detinova curve”, \( EIP = \frac{111}{T-16} \) (Detinova, 1962; Patz and Olson, 2006). Since the EIP tends to decrease as temperature rises, this supports the argument made by some that the malaria burden may worsen as global temperatures increase (Githeko et al., 2000; Patz and Olson, 2006; Paaijmans, Read, and Thomas, 2009). The interaction between lifespan and EIP in expression (1.1) means that, unsurprisingly, any intervention that targets killing adult mosquitoes will necessarily lower the fraction of individuals who survive the EIP, and can significantly impact disease transmission.

A key assumption underlying the model from which expression (1.1) results, is that mosquito age does not affect the rate at which they die. Since work by Clements and Paterson, 1981 indicated that mosquito populations may experience the effects of old age some have questioned whether this is likely to be the case in reality, and has lead to attempts to detect mosquito senescence. In Chapters 2 and 3, I use meta-analyses of MRR experiments and dissection-based studies, respectively, to determine whether mosquitoes experience the effects of old age. To motivate this aim, I now summarise the current modelling approaches that have been used to assess the importance of mosquito senescence for malaria transmission, and also the current experimental evidence for this aspect of mosquito ecology.
1.8 The importance of mosquito senescence

In the Appendix to this thesis, I include a brief toy model of form similar to that presented by Bellan, 2010. This modelling illustrates that age-dependent mortality can affect the impact of vector control programmes, however, unsurprisingly, the importance of this facet of mosquito ecology depends on the degree of the effect strength in wild populations. The results that I present in Chapters 2 and 3 suggest that the magnitude of age-dependent mortality faced by wild populations is relatively minor and, hence, I argue that the impact of mosquito senescence is more likely to be marginal rather than revelatory. I also believe that this assertion is supported by the results of other attempts to model this effect.

Novoseltsev et al., 2012 model the impact of different mortality curves on vectorial capacity and $R_0$, finding that the impact of age-dependent mortality is most important when vector densities and/or population growth rates are sufficiently low, so that $R_0 \sim 1$. In this knife-edge case, the form of age-dependent mortality assumed can determine whether or not enough mosquitoes survive past the EIP to allow a mosquito borne pathogen to persist.

Hancock, Thomas, and Godfray, 2009 model the use of fungal biopesticides, which kill malaria-carrying mosquitoes before they are old enough to transmit the disease, using an age-structured susceptible-exposed-infectious model of adult mosquito lifetime, with a constant rate of larval recruitment. They investigate how different pesticides, each of which inflicts a different age-related mortality risk (modelled using a Weibull distribution) on a mosquito population, affect the transmission of disease. This modelling suggests that the shape of the curve of mosquito mortality is an important factor for determining the impact of spraying on disease transmission. In their simulations, disease transmission was most sensitive to the shape of the additional mortality hazard when considering highly effective pesticides. The age-dependent action of these pesticides is, however, likely more extreme.
than the natural mortality faced by mosquitoes (those populations in the absence of vector control), meaning that it is possible that the effect of senescence on disease transmission for most populations is more subtle.

A range of physiological mechanisms have been suggested may lead to age-related death, including changes in mosquito flight (Nayar and Sauerman, 1973), immune functioning (Christensen, LaFond, and Christensen, 1986), and detoxification mechanisms (Hazelton and Lang, 1984; Lines and Nassor, 1991). To my knowledge, however, experiments have not been done that demonstrate which, if any, of these factors influence wild populations.

A number of studies have also found evidence for mosquito senescence, in both lab populations (Styer et al., 2007; Dawes et al., 2009), and also for wild mosquitoes (Clements and Paterson, 1981; Harrington et al., 2008; Hugo et al., 2014). In large cage experiments consisting of over 100,000 *Ae. aegypti* mosquitoes, Styer et al., 2007 found that the mortality rate increased with age across both mosquito sexes, where females were found to live about twice as long as males (females had a mean lifespan of c.32 days whereas for males it was c.16 days). Similarly, Dawes et al., 2009 conducted experiments on 1,000 *A. stephensi* (an important vector of malaria in India) and concluded that mortality depended on age. In both cases, the mortality rate was low and relatively constant, however, for mosquitoes younger than 10 days old. Since estimates of wild mosquito lifespan are often less than 10 days (see Chapters 2 and 3), however, it is less likely that senescence will be as important a factor for wild populations since most mosquitoes will die due to extrinsic causes before intrinsic mortality affects them. Additionally, Dawes et al., 2009 found evidence that infection with *P. berghei* reduced the lifespan of mosquitoes. Whilst this result may hold for wild populations, it is unlikely given the millions of years that the parasite has probably had to adapt to minimise its impact on the mosquito host (Carter and Mendis, 2002) and, instead, may reflect the inevitable inbred nature of laboratory colonies. More generally,
this suggests that age-related mortality may be a more significant factor in laboratory colonies than it is for wild mosquitoes.

In a meta-analysis of MRR experiments and mosquito dissection studies conducted on eleven different species of mosquito, Clements and Paterson, 1981 compared the fit of an exponential distribution for lifespan with that of a Gompertz model. In a majority of experiments (mainly corresponding to dissection based studies), the Gompertz model - representing an increase in mortality with age - provided a better fit to the data. Harrington et al., 2008 performed MRR experiments on cohorts of laboratory-reared *Ae. aegypti* mosquitoes, consisting of ages between 0 and 20 days, and found that the number of recaptured individuals decreased with the age of the cohort; suggesting an elevated mortality for older mosquitoes. Because mosquitoes in the laboratory probably live artificially long, however, it is possible that the older cohorts were already experiencing the effects of age-related decay before release and, therefore, these results may not generalise to wild populations. Hugo et al., 2014 used a novel technique of mosquito ageing, based on transcriptional profiling of individual mosquito specimens, to estimate *Ae. aegypti* mosquito mortality across different times of the year in Vietnam. In the wetter and cooler of the seasons, mosquitoes were estimated to live shorter lives, and the exponential model - which assumes an age-independent mortality risk - was a better fit to the data. In the dry and hot season, the mosquitoes lived longer, and the Gompertz distribution provided a better fit to the data. This mixed result suggests that mosquito mortality may be more complex than is currently believed, and outside factors, such as the weather and the availability of hosts, may affect the pattern of mortality.

Overall, I argue that on the basis of previously-published work, it is unclear whether age-related mosquito mortality matters for wild populations. This, in part, probably reflects the subtlety of the effect, and the cost of carrying out experiments with the large number
of mosquitoes needed to detect it. It also motivates the approach I take in Chapters 2 and 3, where I pool data from a number of studies to estimate mosquito mortality.

1.9 Future vector control by genetic drive

So-called ‘selfish genes’ (Dawkins, 1976) can spread through a population even when they confer a fitness cost to the individual. The eminent evolutionary biologist Bill Hamilton (Hamilton, 1967) recognised the significance of these types of genes for controlling pests. In a landmark paper, Hamilton argued that a gene on the Y chromosome in a heterogametic male that causes this chromosome to beat the X chromosome in a race to fertilise the egg, would have a selective advantage. As the mutant gene spreads, it would drive sex ratios away from the 1:1 equilibrium argued for by Fisher, 1930, resulting in smaller and smaller populations that are increasingly male-biased. Eventually, stochastic loss of the few remaining females would then lead to eradication of the local population. The idea of using gene drive for vector control (Hickey and Craig Jr, 1966; Hamilton, 1967) and pest management (Curtis, 1968) was first proposed in the 1960s, yet serious attempts to introduce gene drive systems into organisms did not occur until the twenty first century.

Homing endonuclease genes (HEGs) code for nonessential genes which occur in a variety of unicellular organisms, including yeast, bacteria, fungi and viruses (Mueller et al., 1993; Belfort and Roberts, 1997). HEGs code for a protein that targets a c.24 base pair sequence in the DNA and causes a double-strand cut there (Stoddard, 2005). For heterozygous individuals, the cell typically repairs the break using the chromosome containing the HEG as a template and, hence, the HEG is copied over to the other chromosome, in a process known as homing, converting a heterozygote into a homozygote (Stoddard, 2005). This trick allows HEGs to spread from rare to fixation in populations. Given the selective advantage that these genetic elements possess, it appears surprising that they do not occur
in larger multicellular organisms. This is because, after reaching fixation, the absence of further selection means that these elements can accumulate mutations and become non-functional (Godfray, North, and Burt, 2017). Indeed, Goddard and Burt, 1999 found evidence in yeast that horizontal transmissions of HEGs between species are required for long-term persistence of these genes.

Burt, 2003 was the first to propose that HEGs could be used to control mosquito vector populations. Burt imagined a number of different ways a HEG could be used to impose a fitness load on either the pathogen population and/or the mosquito population. First, a HEG could be engineered to recognise a site which is essential for the pathogen yet non-essential for the mosquito. The gene would spread to fixation and, in doing so, replace the target gene, meaning that the pathogen is unable to be transmitted by the mosquito. Second, a HEG could target a gene that is essential for the fitness of the vector, for example, a gene associated with female fecundity or survival. If the HEG spread to fixation, it would impose a large a genetic load on the mosquito population. This would likely lead to population suppression, potentially to a level where the pathogen could be lost from circulation. Third, Burt imagined that a HEG could be inserted onto the Y chromosome but recognise a site on the X chromosome, under a promoter that caused it to be expressed only at male meiosis. When the gene was expressed, it would cause the X chromosome to be cut, and attempts to repair it would result in a non-viable X gametes, meaning all of the offspring would be sons (Godfray, 2013).

The recent discovery of the Crispr-Cas9 system in bacteria (Jinek et al., 2012; Doudna and Charpentier, 2014), and the idea that it might be used to introduce artificial homing mechanisms into wild populations (Esvelt et al., 2014; Simoni et al., 2014) has provided further stimulus for research into the use of driving endonuclease gene systems (DEGs) - a term proposed by Godfray, North, and Burt, 2017 for the collection of constructs
which have characteristics similar to HEGs - to control vector populations. Mathematical
modelling of DEGs has played a crucial role in shaping the research into this area, and the
field is expansive (see Godfray, North, and Burt, 2017 for a summary of the work done,
thus far). Rather than attempt to summarise all models that have been proposed, I discuss
a few indicative cases.

Many current proposals for DEGs aim to cause harm to the mosquito vector population.
Whether a driving endonuclease gene (DEG) will spread is a function of the fitness cost
(s) it imposes on a homozygous individual (for this analysis, I assume that there is no cost
for carrying a DEG on a single chromosome), and the rate at which homing occurs (e), in
other words, the rate at which a heterozygote is converted into a homozygote. Following the
derivation in Godfray, North, and Burt, 2017, I use a non-spatial model to determine the
equilibrium frequency reached by the DEG $q$, where equal numbers of copies produced for
randomly chosen wild-type and DEG alleles should be equal. A randomly chosen wild-type
allele will be present in a wild-type homozygote with probability $(1 - q)$ and be transmitted
to half the offspring, and will be present in a heterozygote with probability $q$, where the
probability it is transmitted to its offspring is $\frac{1}{2}(1 - e)$. A randomly chosen DEG allele will
be present in a DEG homozygote with probability $q$, where it will be present in half of the
offspring that survive $(1 - s)$, and in a heterozygote with probability $(1 - q)$, where it be
present in $\frac{1}{2}(1 + e)$ of the offspring. Equating the numbers of offspring for wild-type and
DEGs, we obtain,

$$\frac{1}{2}(1 - q) + \frac{1}{2}q(1 - e) = \frac{1}{2}q(1 - s) + \frac{1}{2}(1 - q)(1 + e),$$  \hspace{1cm} (1.2)

which, when rearranged yields,

$$-qe = (1 - q)e - qs,$$  \hspace{1cm} (1.3)
which can be interpreted intuitively as the net effect of a DEG on a wild-type allele (left-hand side) must equal its net effect on another DEG allele (right-hand side). I now solve for the equilibrium frequency of a DEG using expression (1.3), to obtain,

\[ q = \frac{e}{s}. \]  

If \( e < s \), a stable polymorphism with both alleles present will exist in the population. If \( e > s \), meaning that the rate at which homing occurs exceeds the fitness cost of a homozygous DEG, then \( q > 1 \) (which is impossible), indicating that the DEG will spread to fixation and the wild-type will disappear.

The load that a DEG imposes on a population at equilibrium is straightforward to calculate. At equilibrium, the frequency of homozygous DEGs equals \( q^2 \), meaning that the relative population fitness decreases from 1 to \( 1 - q^2/s \). Using expression (1.4), the relative fitness of a population at equilibrium hence equals \( 1 - e^2/s \). As \( e \to 1 \), then DEGs targeting genes essential for survival (\( s = 1 \)) would spread and, therefore, the overall fitness decreases to 0, and population elimination is likely.

Because the action of a DEG interacts with other processes that regulate the population, it is difficult to determine exactly what its impact will be on population size. Of particular importance is to consider how a DEG might interact with processes which are density dependent (Alphey and Bonsall, 2014; Godfray, North, and Burt, 2017), because these are the processes that typically regulate the population size. In the mosquito vectors responsible for the bulk burden of disease, it is generally believed that density dependence acts only in the larval stage (Gimmig et al., 2002; Muriu et al., 2013). This means that it matters when during their life the fitness cost resulting from the DEG’s action is paid. If the DEG acts during the egg stage, for example, this could lower competition for resources amongst larvae, which might result in similar numbers of individuals surviving the
larval stage, with little change to the adult population. Ideally, for mosquitoes, a DEG would, hence, act at the pupal stage or on newly emergent adults, to avoid the interaction with density dependence (Godfray, North, and Burt, 2017). It has also been shown that the precise shape of the function dictating density dependent mortality can affect population dynamics (May, 2001). For example, were density dependence to act in an over-compensating way - where low population densities in one generation lead to much larger densities in the next, and *vice versa* - then this would certainly be of concern to a DEG release (Alphey and Bonsall, 2014). However, Godfray, North, and Burt, 2017 argue that this sort of population regulation probably does not occur in mosquito populations and, hence, is not likely to influence population dynamics of an introduced DEG.

The DEG that I have considered, thus far, is assumed to knock out a gene that affects the fitness of an individual. As I have already discussed, however, other varieties of DEGs are also conceivable. Deredec, Burt, and Godfray, 2008 considered a HEG inserted on the Y chromosome which causes cuts in the X chromosome, rendering it non-viable, sometimes called an *X-shredding* or *Y drive* construct. Using a non-spatial population genetic model, they demonstrate that the equilibrium sex ratio (the proportion of males) is given by 

\[ \frac{1}{1 + (1 - e)^k} \]

where \( e \) is the rate of homing and \( k \) is the number of HEGs (each with a unique target site on the X chromosome) inserted onto the Y chromosome. As the rate of homing is increased, therefore, the sex ratio increases towards 1, meaning a fully male population is obtained. The more independent HEGs that are inserted onto the Y chromosome, the greater the effective rate at which homing occurs, meaning that a male-biased population can be achieved at lower homing rates for a single construct.

As I have already discussed, the malaria eradication campaign initiated in 1955 largely overlooked the difficulty of eradicating malaria in sub-Saharan Africa, where access can be problematic and the public health infrastructure is generally poor. Today, the condition of a
number of countries in the region (for example, the Democratic Republic of Congo, Burkina Faso and Somalia) means that the gains from the recent declines in disease transmission have been distributed unequally throughout the continent (Bhatt et al., 2015). A benefit of a gene drive campaign is the ability of a DEG to spread across a region (being carried by the insects, themselves), meaning that it likely needs less on the ground support than some other interventions. The spatial spread of DEGs has been considered in a number of mathematical models, one of which I extend in Chapter 5. In the same Chapter, I discuss implications of spatial models describing the population dynamics of a Y drive HEG and, hence, do not discuss this aspect further now.

Until now, all the vector control measures that have been widely deployed against mosquitoes and the malarial pathogens have suffered as a result of evolved resistance. It is thus pertinent to consider how resistance might arise in response to the release of a DEG, and how this may influence the resultant spread of disease. The most likely form of resistance allele to arise is one that does not contain the DEG recognition site (Godfray, North, and Burt, 2017). Whilst considerable efforts are under way to locate recognition sites that are highly conserved, it is possible that a few wild-type forms not containing the recognition site may naturally exist. Alternatively, the homing action itself could cause a mutation in the recognition site (Godfray, North, and Burt, 2017). Homing typically results when a double-strand break in the DNA is repaired using the homologous chromosome as a template. There are, however, non-homologous ways in which the DNA can be repaired. One particular mechanism is called end joining (EJ) and involves the ligation of the loose chromosome ends (Weterings and Chen, 2008). If the DEG recognition site is repaired during this process, EJ simply results in a reduction of homing efficiency. However, if EJ destroys or significantly mutants the recognition site, then a mutant allele will arise which confers resistance to the homing. If the mutant allele has a fitness similar to the one
with the recognition site, then modelling suggests it will spread throughout the population (Derdec, Burt, and Godfray, 2008). Perhaps the most straightforward way to mitigate against this risk is to engineer a DEG to cut at a number of sites in the target gene (Alphey, 2016; Godfray, North, and Burt, 2017), or to combine multiple HEGs in a single intervention (Godfray, North, and Burt, 2017).

It is now an exciting time to work in vector biology since a number of laboratories have successfully introduced DEGs into mosquitoes (Windbichler et al., 2011; Gantz et al., 2015; Hammond et al., 2016), and work on the technology is continuing at a fevered pace. There is also a research consortium called Target Malaria (of which I am part), whose aim is to consider these sorts of technology for use in malaria control and, when the crucial approval from stakeholders, regulators and fellow scientists is obtained, hopefully to use it to help control the disease.

1.10 Motivating this thesis

The impact of mosquito-borne disease on historical cultures can be read about in books ranging from literature to archaeology. The current impact of mosquito-borne disease on some developing countries can, however, be seen first hand. History teaches us that a single strategy for controlling, and later, eradicating these diseases, will likely fail, and that multi-pronged approaches are necessary. Vector control has been a critical factor in those successful past attempts to eradicate vector borne disease from a given location, as well as the recent declining incidence of malaria. Yet, mosquito resistance to insecticides, climate change, and donor fatigue, amongst other factors, threatens to weaken the potency of these methods. In this context, it becomes even more crucial to better understand mosquito ecology, in order to assess existing interventions, and develop new policies which effectively eliminate disease. A crucial element of mosquito ecology, to which the majority
of this thesis is devoted, is the length of time that mosquitoes live. In Chapters 2 and 3, I use data from the two prominent approaches to estimate this quantity - MRR studies and dissection-based experiments - to derive estimates of mosquito longevity which reflect the weight of data that has been collected. In Chapter 4, I analyse the results of experiments from a new technology, known as near-infrared spectroscopy, which can age individual mosquitoes, and show that it may be worth pursuing as a tool to estimate mosquito lifespan in the wild.

The worrying recent parasite resistance to insecticide, which has been recorded in Asia, means that now, more than ever, it is essential for us to develop new methods for disease control. In Chapter 5, I model a recently-proposed method for vector control, using a HEG which biases the sex ratio towards males. If such a gene drive mechanism were released into the wild, it will be important to monitor its effects on the mosquito population, particularly to search for any signs of resistance. It is conceivable that the three types of method which I have mentioned, MRRs, mosquito dissection, and near-infrared spectroscopy, could be used for this purpose.
Chapter 2

A meta-analysis of mark-release-recapture experiments of mosquitoes

2.1 Abstract

Mosquito-borne disease transmission is sensitive to the duration of adult mosquito lifespan. An important method to estimate lifespan in the field is mark-release-recapture (MRR) experiments. These experiments, however, are costly to undertake and can resort in estimates associated with considerable uncertainty. Also, there is considerable heterogeneity in the methods used to analyse such experiments, meaning that it is not straightforward to compare estimates. In this chapter, I apply a common statistical methodology to the analysis of a dataset of 232 historical MRR experiments, which allows direct comparison between the estimates. Further, I use a Bayesian hierarchical model to produce estimates of lifespan at the species and genus levels and, also, estimates of lifespan by sex. This
analysis indicated that there were significant differences in the lifespan of mosquitoes between species and genera. The analysis indicates, also, that female mosquitoes live on average for 0.8 days longer than males. By fitting a range of different survival models, each of which embodies different assumptions about age-dependent mortality, I test whether mosquitoes experience senescence. In all cases, the exponential survival model, which assumes a mortality risk which is independent of age, fit the data as well as or better than each of the five other models which incorporated age-dependent mortality, suggesting that wild populations of mosquitoes do not undergo senescence.

2.2 Introduction

Of particular importance to the epidemiology of mosquito-borne disease is the length of time that adult mosquitoes live in the wild, which has a fourfold effect on disease transmission. All else being equal, mosquitoes with longer lifespans (i) form larger populations, (ii) are more likely to become infected, (iii) are more likely to survive the extrinsic incubation period (EIP) of the disease to become infectious and (iv), after becoming infectious, are more likely to bite and infect a human (Ross, 1910; Macdonald, 1957; Anderson and May, 1992; Smith and McKenzie, 2004). The sensitivity of disease transmission to these factors means that information on mosquito longevity is crucial to assess disease control strategies.

There are currently two methods that have been used to determine mosquito lifespan in the wild: MRR studies and dissection-based approaches that estimate the age of captured samples using a knowledge of the physiological effects of ageing on mosquitoes. The first MRR experiments were used to study the flying distance of anopheline mosquitoes in the Panama during construction of the canal (Zetek, 1915) and in South Carolina, USA (Le Prince and Griffitts, 1917) but, subsequently, these experiments have been employed to determine lifespan (see Guerra et al., 2014 for a summary of MRR experiments). In MRR
experiments mosquitoes are marked (typically with a fluorescent die) and released into the
wild, and subsequent captures of wild mosquitoes are carried out over a number of days.
Over time the released mosquitoes die and thus the number of marked mosquitoes that
are recaptured declines. The faster the rate of decline the higher the mortality hazard,
and the shorter the lifespan. Hence by monitoring the decline in recaptures over time the
lifespan of the marked mosquitoes can be estimated by statistical methods. Dissection-
based methods have mostly been only applicable to female mosquitoes, and relies on the
assumption that irreversible changes occur to female reproduction organs throughout their
lifetime (Polovodova, 1949; Detinova, 1962). These methods either examine the proportion
of mosquitoes that have laid eggs, or estimate the number of gonotrophic cycles a female
mosquito has experienced by counting the number of beadlike dilations in the ovarioles
(Polovodova, 1949; Gillies and Wilkes, 1965; Samarawickrema, 1968).

Dissection-based methods are used less frequently than MRR experiments to determine
mosquito lifespan. We speculate on the reasons for the lower popularity of this method:
MRR experiments can be used to estimate other characteristics of wild mosquito popu-
lations, such as mosquito dispersal and population sizes; dissecting individual mosquitoes
requires specialised knowledge and equipment; and dissection-based approaches require
the experimenter to make assumptions about the effect of the gonotrophic cycle on fe-
male mosquito physiology. MRR experiments require fewer assumptions to be made about
mosquito biology to arrive at estimates of lifespan (Reiter, 2007): behaviour and survival
are not affected by marking or handling mosquitoes; the probability of recapture does not
change with age; and the rate of mosquito dispersal is small compared to the risk of mortal-
ity. The assumption that MRR estimates of lifespan are unaffected by mosquito dispersal
is likely violated in practice, and we believe strengthens the need for other methods (such
as dissection-based approaches) that do not depend on mosquito dispersal.
Unfortunately, individual MRR studies are costly and can result in poor quality estimates of mosquito lifespan, due to the difficulty of capturing mosquitoes in natural environments. Furthermore, the variety of experimental and statistical methods used across different studies makes it difficult to reach a consensus about the lifespan of wild mosquitoes.

In this chapter, we overcome some of these issues by combining data from 232 separate MRR time-series covering 33 separate species, and three genera, from a recently-published database (Guerra et al., 2014). We apply a hierarchical Bayesian model to these data, allowing us to estimate average lifespans under a range of assumptions about how the individual time-series are related. In particular, we estimate lifespan at both the species and genus levels. We also examine the effect of exogenous factors such as air temperature, mosquito sex, and blood-feeding.

In addition to our investigations of average mosquito lifespan, we further ask if there is evidence that mosquito mortality increases with age in wild populations. This is of interest because a mosquito population whose members are subject to senescence will contain a smaller fraction of very old mosquitoes, which are particularly effective in transmitting disease, than a similar population where senescence does not occur (Styer et al., 2007; Bellan, 2010; Novoseltsev et al., 2012). To address this question we use six different statistical models of mosquito mortality taken from mosquito studies. Each of these models embodies different assumptions about senescence and we fit the parameters within our Bayesian framework.
2.3 Materials and methods

2.3.1 Data

The MRR database in Guerra et al., 2014 contains 393 individual time-series, along with meta-data for a range of factors for each experiment (for example, species, geography, and date of study). We then transformed the data into a form amenable to our statistical model. In particular, we removed time-series with fewer than six separate recapture observations, as well as species with fewer than two separate MRR time-series. The final data encompassed 232 time-series (see Section 7.2 for a list of the studies included in our final dataset), with a range of characteristics (Tables 2.1 and 2.2). The data encompassed time-series from MRR experiments across 33 different species and three genera: *Aedes* (91 separate time-series), *Anopheles* (94), and *Culex* (47), and spanning a wide geographical range (Figure 2.1). The data comprise 177 female-only, 35 male-only and 18 mixed-sex release time-series. Where the age of the adult mosquitoes at release-time was available (102 time-series) we include this information in our time variable (that represents mosquito age) in the statistical models. Where age was unavailable (130 time-series), we assume that the age at release is zero. We recognise this assumption will likely underestimate mosquito age for some studies and so we interpret our estimates as lower bounds on mosquito lifespan.

2.3.2 Statistical model

Data for a typical MRR experiment consists of a single release of $N_R$ marked mosquitoes followed by a series of marked mosquito recaptures on subsequent days (Figure 2.2 shows three such example series.) We model the number $y(t)$ of marked mosquitoes recaptured on day $t$ using a negative binomial sampling model,

$$y(t) \sim \text{NB}((N_R - Y(t-1))S(t)\psi, \kappa), \quad (2.1)$$
### Data characteristic

<table>
<thead>
<tr>
<th>Data characteristic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of species</td>
<td>33</td>
</tr>
<tr>
<td>Number of genera</td>
<td>3</td>
</tr>
<tr>
<td>Number of <em>Anopheles</em></td>
<td>94</td>
</tr>
<tr>
<td>Number of <em>Aedes</em></td>
<td>91</td>
</tr>
<tr>
<td>Number of <em>Culex</em></td>
<td>47</td>
</tr>
<tr>
<td>Number of female-only MRR</td>
<td>179</td>
</tr>
<tr>
<td>Number of male-only MRR</td>
<td>35</td>
</tr>
<tr>
<td>Number of mixed sex MRR</td>
<td>18</td>
</tr>
<tr>
<td>Number of pre-release blood-feeding only MRR</td>
<td>71</td>
</tr>
<tr>
<td>Number of pre-release sugar-feeding only MRR</td>
<td>41</td>
</tr>
<tr>
<td>Number of pre-release both blood- and sugar-feeding MRR</td>
<td>4</td>
</tr>
<tr>
<td>Number of pre-release neither blood- and sugar-feeding MRR</td>
<td>116</td>
</tr>
<tr>
<td>Number of MRR time-series</td>
<td>232</td>
</tr>
</tbody>
</table>

Table 2.1: Summary of variables across all MRR time-series.

### Data characteristic

<table>
<thead>
<tr>
<th>Data characteristic</th>
<th>Min</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study duration, days</td>
<td>6.0</td>
<td>11.8</td>
<td>10.0</td>
<td>71.0</td>
<td>4.7</td>
</tr>
<tr>
<td>Number of days on which collections took place</td>
<td>6.0</td>
<td>10.1</td>
<td>9.0</td>
<td>47</td>
<td>9.1</td>
</tr>
<tr>
<td>Number of separate release days</td>
<td>1.0</td>
<td>1.9</td>
<td>1.0</td>
<td>23</td>
<td>3.0</td>
</tr>
<tr>
<td>Number released</td>
<td>66</td>
<td>4,929</td>
<td>1,297</td>
<td>86,200</td>
<td>12,043</td>
</tr>
<tr>
<td>Number recaptured</td>
<td>2</td>
<td>163</td>
<td>63</td>
<td>4,090</td>
<td>399</td>
</tr>
<tr>
<td>Recapture percentage</td>
<td>0%</td>
<td>8.6%</td>
<td>5.2%</td>
<td>57.1%</td>
<td>10.04%</td>
</tr>
<tr>
<td>Age at release, days</td>
<td>0</td>
<td>1.8</td>
<td>0</td>
<td>13</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Table 2.2: Summary of data from individual MRR time-series.
Figure 2.1: The location and number of separate time-series for the MRR studies included in this analysis. The area of each circle indicates the number of time-series at each study site. The colour shows the genus of the mosquitoes in each study.

where $Y(t - 1)$ is the cumulative number of mosquitoes caught on all days before $t$, $S(t)$ is the probability that an individual mosquito survives and remains in the study area until time $t$, $\psi$ is the daily recapture probability for an individual mosquito which is assumed constant through time, and $\kappa$ is the time-independent shape parameter of the negative binomial distribution that controls the extent of variance in recapture rate likely due mostly to environmental heterogeneity. We chose a parameterisation of the negative binomial such that its mean is given by $\mu(t) = (N_R - Y(t - 1))S(t)\psi$, and its variance by $\sigma(t)^2 = \mu(t) + \frac{\mu(t)^2}{\kappa}$. We use the negative binomial sampling distribution for its ability to account for temporal aggregation in captures due to environmental heterogeneity (Bliss and Fisher, 1953), and has previously been used to model mosquito MRR experiments (Service, 1971; Nedelman, 1983).

For some MRR experiments, there were a number of releases ($q \geq 2$) of marked
mosquitoes throughout the duration of the study. In contrast to single releases, with multiple releases over time, it is not in general possible to determine the particular release to which a recaptured mosquito belongs. To avoid the complication of directly inferring this quantity, we choose to represent previous recaptures probabilistically. This results in a slightly different mean to that of the single release model,

\[ \mu(t) = N_{\text{Released}}(1 - \psi)^{t-1}S(t)\psi, \]  

where the factor of \((1 - \psi)^{t-1}\) represents the probability that a mosquito is not recaptured on a previous day. Where experiments consisted of two or more releases occurring at distinct points in time we assumed that recaptures of individual mosquitoes from either batch were independent of one another, although with the same sampling parameters \((\psi \text{ and } \kappa)\). This results in an overall mean composed of the sum of those from all \(q\) releases,

\[ \mu(t) = \mu_1(t) + \ldots + \mu_q(t). \]  

Figure 2.2: Example MRR time-series for three different studies from the Guerra et al., 2014 database. In all the above experiments there was a single release of a batch of marked mosquitoes.

The simplest assumption for mosquito survival is that death and dispersal rates are
independent of mosquito age, giving an exponential ‘survival’ function,

\[ S(t) = e^{-\lambda t}, \tag{2.4} \]

where \( \lambda \) is the sum of the rates of death and dispersal from the study area. To estimate the lifespan of mosquitoes we use this model because we find no evidence in support of age-dependence across all species (see below for the method used to determine this, and later the ‘Results’).

We assume that the number of mosquitoes recaptured on one day is independent of the number recaptured on any other day, apart from their joint dependence on \( \lambda, \psi \) and \( \kappa \). This results in a likelihood of the data of the form,

\[ L(y(t_1), y(t_2), ..., y(t_R)|\lambda, \psi, \kappa) = \prod_{i=1}^{R} p(y(t_i)|\psi, \lambda, \kappa), \tag{2.5} \]

where \( R \) is the number of individual days where collections took place, and \( p(y(t_i)|\psi, \lambda, \kappa) \) is the probability of recapturing \( y(t_i) \) mosquitoes on day \( t_i \) determined from a negative binomial distribution with parameters \( \mu = e^{-\lambda \psi} \), and \( \kappa \).

### 2.3.3 Individual time-series estimates

We first treat each time-series separately, and estimate individual \((\lambda, \psi, \kappa)\) parameters of the statistical model (described in the previous section) for each time-series. To use a Bayesian methodology we must specify priors on this set of parameters. Here we choose to specify independent priors on each parameter of the form: \( \lambda \sim \mathcal{N}(-1.8, 1) \), \( \psi \sim \exp(20) \) and \( \kappa \sim \text{log-normal}(1, 1) \). This prior on the rate parameter of the exponential distribution \( \lambda \) corresponds to a wide range of possible lifespans (Figure 2.3A), with a mean of 10 days.

It was necessary to use an informative prior on \( \psi \) to allow it to be estimated; the prior on
$\kappa$ is fairly uninformative and allows a wide range of values (Figure 2.3B,C).

![Figure 2.3: The prior probability distributions used for lifespan (A), recapture probability (B) and dispersal parameter (C) in the individual time-series analysis for the MRR studies.](image)

### 2.3.4 Estimating lifespan at the species, genus and overall groupings

We want to synthesise information from across all MRR experiments to produce more robust estimates of mosquito lifespan than can be obtained from considering the individual time-series separately. However there exists considerable heterogeneity across the experiments. This heterogeneity has two sources. There is that arising from variability in experimental methodology. However there is also variability arising from actual differences in lifespan across the different mosquito cohorts; for example due to genetic differences between mosquito populations or due to climatic differences. Because of this heterogeneity we argue that it is not appropriate to assume the parameters of our models are the same across all time-series. We therefore choose to use a Bayesian hierarchical model which is akin to a ‘random effects’ model in classical statistics. This type of model assumes that there is random variation in parameters at the individual time-series level, although each of the parameters is drawn from a common ‘population-level’ distribution. In our case we separately estimate three different model groupings where the population-level distributions correspond to the species, genus and overall (across all studies) levels respectively.
(Figure 2.4). This allows us to estimate average mosquito lifespan for a given species or genus, or alternatively across all time-series, by independent sampling from the posterior predictive distribution from which the individual $\lambda_i$ are drawn.

Figure 2.4: A representation of the hierarchical Bayesian model, for the constant mortality model case when considering the species-level grouping. Here ‘pdf’ indicates ‘probability density function’, and $\lambda_i$, $\psi_i$, and $\kappa_i$ represent the force of mortality for the exponential distribution, daily recapture probability and over-dispersion parameter of the negative binomial distribution for time-series $i$ in the dataset; $\mu_i(t) = \psi_i e^{-\lambda_i t}$ is the mean number of mosquitoes recaptured on day $t$.

Whilst we assume that the variation in mosquito mortality across experiments is random in nature, we recognise that there may exist systematic sources of variation that we do not include in our model. To account for this potential source of bias we examined how the individual estimates of lifespan correlated with experimental covariates (number of traps, range of traps, trapping method), but found no such evidence of systematic variation in
lifespan (data not shown).

A Bayesian framework requires that we incorporate our pre-analysis beliefs into our estimates through the use of priors for all parameters in a particular model. In our hierarchical model we are required to specify prior distributions at two levels of the analysis. The first of these links the individual time-series estimates with the overarching group-level distributions. The parameters of these group-level distributions are then set prior distributions (Figure 2.4). As an example, we suppose that the rate parameter $\lambda_i$ of our exponential survival model in experiment $i$ is drawn from a group-level distribution which we assume to be log-normal,

$$\lambda_i \sim \text{log-normal}(\mu_\lambda, \sigma)$$  \hspace{1cm} (2.6)

where $\mu_\lambda$ and $\sigma$ are the location and scale parameters of the log-normal distribution that summarise group-level characteristics. We then specify priors on these group-level parameters, $\mu_\lambda \sim N(-1.8, 1)$ and $\sigma \sim \text{log-normal}(-2, 1)$. The relatively complex nature of priors in hierarchical models make it important to determine their influence on inferences. We chose the above – somewhat uninformative – priors to allow a range of mosquito lifespans in order to minimise their effect on the estimates we report (Figure 2.5). We also chose to set hierarchical priors on the remaining parameters in our models – $\psi$ the probability of daily recapture, and $\kappa$ the ‘over-dispersion’ parameter of a negative binomial distribution. For $\psi$ we chose a fairly informative prior that placed most weight where $\psi \leq 10\%$, since in most experiments the overall recapture percentage was well beneath this value. For $\kappa$ we chose a fairly wide prior that had most of its probability mass below $\kappa = 20$ (Figure 2.6). In all cases the hierarchical priors were chosen to have similar implications on lifespan, recapture probability and over-dispersion at the individual time-series level as for the non-hierarchical analysis described previously.
2.3. MATERIALS AND METHODS

2.3.5 Testing for age-dependent mortality in wild mosquitoes

Previous work has found evidence of age-dependent mortality in lab populations (Styer et al., 2007; Dawes et al., 2009), and less-commonly in wild mosquitoes (Clements and Paterson, 1981). Furthermore, recent modelling work has examined the implications of departures from a constant risk of mortality (Styer et al., 2007; Hancock, Thomas, and Godfray, 2009; Novoseltsev et al., 2012). To determine whether senescence occurs in wild mosquitoes we re-estimate our model using survival functions, \( S(t) \), that allow for a rate of mortality that can vary with age. Specifically, we re-estimate our model using five other models that were previously used in the literature (see Table 2.3 for a description of these). Each of these models makes different assumptions about how the rate of death and dispersal varies with age, but all can be represented by a general form,

\[
S(t) = e^{- \int_0^t \lambda(\tau) d\tau}.
\]  

(2.7)

where we constrain the parameters of our models to preclude the possibility of a hazard that decreases with age \( \frac{d\lambda(t)}{dt} \geq 0 \). Whilst this could occur if older mosquitoes disperse less than younger mosquitoes, we do not believe this effect would outweigh any declines in survival associated with old age.

The parameters of the survival function of each model were assigned hierarchical priors that allowed considerable variation in mosquito lifespan (see Figure 2.5). Otherwise the statistical model we used remained the same as for the constant mortality case.

2.3.6 Model estimation by MCMC

The likelihood and priors we use result in posterior distributions whose analytic form cannot be calculated with existent computational methods. Instead we use Markov Chain Monte
### Table 2.3: A description of the survival functions used in this study, arranged in rough order of model complexity (simple-complex from top-bottom.)

All parameters are defined to be non-negative. $\Gamma(\theta)$ and $\Gamma(\theta_1, \theta_2)$ refer to the Euler gamma function, and incomplete gamma function respectively. The ‘mean time spent in study area’ is an estimate of the combined effects of mosquito mortality and dispersal from the area of the study where collections take place, since our data does not provide spatially-resolved data.

<table>
<thead>
<tr>
<th>Survival function</th>
<th>Hazard rate</th>
<th>Interpretation</th>
<th>Mean time spent in study area, $\overline{T}$</th>
<th>Papers assuming this function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\lambda$</td>
<td>Constant mortality risk</td>
<td>$\frac{1}{\lambda}$</td>
<td>Ross, 1910; Anderson and May, 1992; Smith and McKenzie, 2004</td>
</tr>
<tr>
<td>Gompertz</td>
<td>$\alpha e^{\beta t}$</td>
<td>Mortality increases with age at an ever-increasing rate</td>
<td>$\frac{e^{\beta t} \Gamma(0, \frac{\beta}{\alpha})}{\beta}$</td>
<td>Clements and Paterson, 1981; Styer et al., 2007; Novoseltsev et al., 2012</td>
</tr>
<tr>
<td>Weibull</td>
<td>$\alpha t^\beta; \beta \geq 1$</td>
<td>Mortality increases with age at an ever-increasing rate</td>
<td>$\alpha^{-\frac{1}{\beta}} \Gamma(1 + \frac{1}{\beta})$</td>
<td>Hancock, Thomas, and Godfray, 2009; and Carey, 2001 considering general insect demography</td>
</tr>
<tr>
<td>Gompertz-Makeham</td>
<td>$\alpha e^{\beta t} + c$</td>
<td>Two additive mortality risks: one that increases with age, and another that is age-independent</td>
<td>No simple analytic form</td>
<td>Styer et al., 2007</td>
</tr>
<tr>
<td>Logistic</td>
<td>$\frac{\alpha e^{\beta t}}{1 + \frac{\alpha e^{\beta t}}{c^{\beta t} - 1}}$</td>
<td>Mortality risk increases with age, although at a declining rate</td>
<td>No simple analytic form</td>
<td>Styer et al., 2007; Novoseltsev et al., 2012</td>
</tr>
<tr>
<td>Logistic-Makeham</td>
<td>$\frac{\alpha e^{\beta t}}{1 + \frac{\alpha e^{\beta t}}{c^{\beta t} - 1}} + c$</td>
<td>Two separate additive mortality risks: an age-dependent risk of the same form as the logistic model, and a constant hazard</td>
<td>No simple analytic form</td>
<td>Styer et al., 2007; Bellan, 2010</td>
</tr>
</tbody>
</table>
Survival function | time-series-level priors | Group-level priors
--- | --- | ---
Exponential | $\lambda \sim \text{log-normal}(\mu_\lambda, \sigma)$ | $\mu_\lambda \sim N(-1.8, 1), \sigma \sim \text{log-normal}(-2, 1)$
Gompertz | $\alpha, \beta \sim \text{log-normal}(\mu_\alpha, 0.2)$ | $\mu_\alpha \sim N(-3, 1), \mu_\beta \sim N(-3, 0.5)$
Weibull | $\alpha, (\beta - 1) \sim \text{log-normal}(\mu_\alpha, 0.2)$ | $\mu_\alpha \sim N(-4.8, 1), \mu_\beta \sim N(-4, 1)$
Gompertz-Makeham | $\alpha, \beta, c \sim \text{log-normal}(\mu_\alpha, 0.2)$ | $\mu_\alpha \sim N(-3, 1), \mu_\beta \sim N(-3, 0.4), \mu_c \sim N(-4.5, 0.5)$
Logistic | $\alpha, \beta, s \sim \text{log-normal}(\mu_\alpha, 0.2)$ | $\mu_\alpha \sim N(-3.5, 0.5), \mu_\beta \sim N(-3.2, 1), \mu_s \sim N(-3, 0.5), \mu_c \sim N(-4, 0.5)$
Logistic-Makeham | $\alpha, \beta, s, c \sim \text{log-normal}(\mu_\alpha, 0.2)$ | $\mu_\alpha \sim N(-3.5, 0.5), \mu_\beta \sim N(-3.2, 1), \mu_s \sim N(-3, 0.5), \mu_c \sim N(-4, 0.5)$

Table 2.4: The priors used on parameters of each different survival model. For the exponential model the ‘group-level’ priors were the same for the genus and ‘overall’ models that were also estimated. The exponential model was the only model that was simple enough to allow the scale parameter of the log-normal ($\sigma$) to be estimated by the data. The notation $\alpha, \beta \sim \text{log-normal}(\mu_\alpha, 0.2)$ means that $\alpha$ and $\beta$ were assigned independent log-normal priors with location parameters $\mu_\alpha$ and $\mu_\beta$ respectively, and a scale parameter of 0.2 in both cases.

To judge convergence of the sampling algorithm to the posterior density, we calculated $\hat{R}$ across all Markov chains – a ratio that compares the between-chain variation to that within each chain that is commonly used to measure convergence in MCMC (Gelman and Rubin, 1992). For each model that we estimated we conducted an independent MCMC run consisting of 12 independent Markov chains with 1,000 iterations per chain, discarding the first half of these iterations as warm-up (Gelman et al., 2014), and continued to run the algorithm until $\hat{R} < 1.1$ for all model parameters. We also ensured that across each MCMC run the number of divergent iterations (that can bias the MCMC away from the true posterior density) was minimal. In the majority of cases, the number of divergent iterations was far fewer than 1% of the total number of samples.
2.3.7 K-Fold cross validation

One way to estimate a model's out-of-sample predictive capability is to use Akaike Information Criterion, Bayesian Information Criterion, Deviance Information Criterion or Watanabe Akaike Information Criterion that explicitly penalise a model in accordance to its complexity (typically indexed by its number of free parameters) to correct for model over-fit. The way in which this correction takes place however, is fairly heuristic, and less appropriate for hierarchical models. An alternative method, common in the machine learning literature, is known as 'cross-validation' (see for example, Kohavi, 1995). Here to estimate out of sample predictive capability, the original data set is partitioned into training and test sets. The model is then fitted to the training set, and its predictive performance measured on the independent test set.

We use cross-validation to compare the predictive power of the species, genus and overall models, and later to determine whether age-dependent mortality occurs. In particular, we use K-Fold cross-validation where we repeatedly partition our dataset into a test set composed of a number of individual time-series, and a training set of the remainder (see for example, Marsland, 2015). We then use the fitted model to measure the predictive performance on the test set. The particular measure we use is the expected log point-wise predictive density (elpd) (Vehtari, Gelman, and Gabry, 2015) which sums the predictive performance for each data point averaged across all posterior samples.

Since the models are hierarchical we must specify how to draw parameters for a given test sample time-series. Here we chose to draw values of the individual time-series parameters as samples from the population distribution that corresponds to that particular sample's respective grouping. So if the particular test sample time-series pertains to *Ae. aegypti* (and we are using the species-level model) then we draw independent samples for
the statistical model’s parameters from the estimated *Ae. aegypti* population distribution,

\[ \theta_i \sim p(\theta_{Ae. aegypti}|data), \]  

(2.8)

where \( \theta_i \) is the value of the parameters used in test time-series \( i \), and \( p(\theta_{Ae. aegypti}|data) \) is the overall posterior distribution across all *Ae. aegypti* experiments.
Figure 2.5: Samples from the prior distribution of mean lifespan across the six models of mosquito mortality introduced in Table 2.3. In all cases 10,000 samples were generated using the priors indicated in Table 2.4.
Figure 2.6: **Samples from the priors for daily recapture probability** ($\psi$) and **over-dispersion parameter** ($\kappa$). In all cases 10,000 samples were generated using the following priors: $\psi \sim \exp(\theta_\psi)$ and $\kappa \sim \exp(\theta_\kappa)$. 
2.4 Results

2.4.1 Mosquito lifespan varies across species although variation is less marked at the genus level

We first present estimates of mean mosquito lifespan obtained by separate analysis of each individual time-series (Figure 2.7). Whilst there is variability in these estimates, in 211 of the 230 cases we report the median lifespan was below 10 days. (We do not present estimates for the two time-series that corresponded to experiments done on *Anopheles balabacensis* since this data was highly unusual – recaptures increased over time – and resulted in unrealistic estimates.) However, from this separate treatment of individual MRR time-series it is not immediately clear whether differences exist at the species or genus level. To determine this we estimated hierarchical Bayesian models where we grouped by species, genus or across all time-series (‘overall’). These estimates show considerable variation at the species level (Figure 2.8 main), with the lowest estimated mean lifespan corresponding to *Anopheles (A.) subpictus s.l.*, a major vector of malaria in Asia (Sinka et al., 2012), with a median lifespan of 1.0 days. It is unlikely that this estimate is representative of the true lifespan of this mosquito species, since with such a short lifespan there would be negligible transmission of malaria. Since we cannot disentangle the effects of mosquito mortality from dispersal in our estimates, the estimates we present represent lower bounds on mosquito lifespan, which in some cases are very conservative. The highest mean lifespan was estimated for *Aedes (Ae.) cantans*, a mosquito species found in Western Europe (Lakhani, 1974; Hubálek, 2008), with a median lifespan of 14.5 days. ANOVA tests using the posterior medians from each time-series suggested that differences in mean lifespan exist at the species level ($p < 0.01$, $F = 10.76$, $df = 32$), and was robust to non-inclusion of *Anopheles balabacensis* ($p < 0.01$, $F = 2.74$, $df = 31$). The Kruskal-Wallis test also indicated that
this result was robust to non-normal residuals ($p < 0.01$, $\chi^2 = 99.83$, $df = 32$). Lifespan estimates were less variable at the genus level (Figure 2.8 inset): *Culex* mosquitoes had the lowest median estimate at 2.4 days; *Anopheles* were second highest at 3.6 days; and *Aedes* were estimated to have the highest mean lifespan at 4.5 days. ANOVA tests suggested that there were differences in mean lifespan at the genus level ($p < 0.01$, $F = 4.74$, $df = 2$).

Again this result was robust to the assumption of non-normal residuals with the Kruskal-Wallis test results indicating statistically different lifespans ($p < 0.01$, $\chi^2 = 25.48$, $df = 2$).

Across all genera the median posterior lifespan was estimated to be 3.7 days (Figure 2.8 inset).

To determine whether grouping by species explained more of the variation in the data than genus we conducted K-Fold cross-validation (see ‘Methods’) on each of the models. These suggested that the species and genus groupings performed better than the model where we simply pooled all data together ($p_{\text{species}} = 0.02$, $p_{\text{genus}} = 0.04$). However the species model did not lead to better prediction than the one where we grouped by genus, suggesting that the majority of variation in lifespan is explained by genus (Figure 2.9).

Why do we see such variation at the species level, and yet this model is not significantly better at prediction than the genus one? Whilst for some species grouping at this level leads to significantly better predictions, in other cases the model is over-fit to the data and so does not perform as well (Figure 2.10). In biological terms the species-grouping model performs better when the data are representative of real underlying differences in lifespan, but less well when the experimental data are influenced by significant idiosyncratic ‘random’ factors that lead to differences in lifespan estimates.

In an MRR experiment there are two mechanisms that can cause a reduction in mosquito numbers after release: 1. mosquitoes die before they are recaptured, or 2. mosquitoes disperse out of the study area. Since our analysis is based on spatially-averaged
CHAPTER 2. A META-ANALYSIS OF MRR EXPERIMENTS

MRR data our estimates actually represent the combined effect of mortality and dispersal (‘the time in study area’). However, if mosquito dispersal causes a considerable reduction in mosquito numbers after release then studies with a wider recapture radius should have a higher estimated longevity than studies with a smaller radius. The spatial extent of the study should therefore be positively correlated with our estimates if dispersal is an important omitted variable in our models. However there is no such correlation (Figure 2.11).

This lack of correlation suggests that our estimates of lifespan may not be strongly biased downwards. However, it is possible that such a correlation exists due to low recapture success of outer traps in studies with a larger recapture range.

2.4.2 A small proportion of mosquitoes are likely responsible for the bulk burden of disease for malaria and Zika

Whilst our principal estimates of adult mosquito lifespan (Figure 2.7 and 2.8) represent means, it is the distribution of lifespans at the individual level that dictates vectorial capacity. We therefore used the posterior parameter samples for the rate parameter in our exponential distribution to sample quantiles from the survival curves. These curves show the proportion of individual mosquitoes that live up until a given age. In particular, we estimated the proportion of mosquitoes that survived the extrinsic incubation period (EIP) for those species that are identified as vectors for malaria, chikungunya, dengue fever and Zika.

For the Anopheles species included in this analysis there was a large range in the percentage of mosquitoes surviving the EIP for malaria (Figure 2.12) which we assumed to be 10 days (Blanford et al., 2013): our posterior median estimates indicate that <0.1% of An. subpictus s.l. and 46% An. sergenti survived long enough to become infectious with malaria. Further ANOVA tests on the posterior median proportions for the individual
time-series suggested that differences exist in the fraction of mosquitoes that survive the EIP for malaria \((p < 0.01, F = 32.91, df = 16, n = 88\) time-series). The Kruskal-Wallis test confirmed this result was robust to non-normal residuals \((p < 0.01, \chi^2 = 62.43, df = 16, n = 88\) time-series). Importantly, for the major African vectors of malaria, \textit{An. gambiae} \textit{s.l.} and \textit{An. funestus} \textit{s.l.}, we obtained posterior median estimates that indicate that only 2% and 4% of individual mosquitoes survive the EIP.

For the three other diseases we consider: chikungunya, dengue fever and Zika, the two species \textit{Ae. aegypti} and \textit{Ae. albopictus} are the vectors responsible for causing the majority of disease cases (Pialoux et al., 2007b; World Health Organisation, 2002; Bogoch et al., 2016). The EIP for Chikungunya is the shortest of the diseases we consider, at around 2 days (Dubrulle et al., 2009). The proportion of mosquitoes surviving this period is considerably higher than for malaria (Figure 2.12, bottom): we estimate that 59\% \textit{Ae. aegypti} and 76\% \textit{A. albopictus} mosquitoes live long enough to transmit chikungunya. The EIP for dengue fever is thought to depend particularly strongly on temperature, and so we consider the EIP at a fixed temperature of 30\(^\circ\) C, estimated to be approximately 6.5 days in a recent meta-analysis (Chan and Johansson, 2012). Assuming this EIP we estimate that 22\% \textit{Ae. aegypti} and 42\% \textit{Ae. albopictus} mosquitoes live long enough to pass on dengue fever. Whilst the research into Zika is at a fairly early stage, the existent estimates of the EIP for the disease suggest that its EIP is longer than for the other diseases we consider. Here we assume an EIP of 12.5 days, which represents an average of two recently-published estimates (Chouin-Carneiro et al., 2016; Di Luca et al., 2016). For this EIP we estimate that 8\% \textit{Ae. aegypti} and 19\% \textit{A. albopictus} survive to be able to transmit Zika.
2.4.3 The longevity of mosquitoes does not vary systematically over the range of temperatures of MRR study locations

The Guerra et al., 2014 database included the latitude and longitude of each study, along with the date when the study began. We used this information to find estimates of the air temperature for each study using the European Centre for Medium Range Weather Forecasts’ historical database. For each study we calculated the mean monthly temperature across a spatial area of (lat ± 1 degree, long ± 1 degree), for the month at which each study was carried out. Unfortunately, the records for this database begin in 1979, which pre-dates the study date for 65 of our 232 MRR time-series. For these time-series, we chose to estimate the air temperature by an average of monthly temperatures over the years 1979-89.

We carried out a regression of the posterior median lifespan estimated for each time-series (using the hierarchical model where we grouped by species) against air temperature including both linear and quadratic temperature terms (Figure 2.13). The effect of temperature was not statistically significant at the 5% level.

2.4.4 Male mosquitoes have shorter lives than females

The Guerra et al., 2014 database contains information on whether the releases consisted only of females, males or a mixture of mosquitoes. For some species in the database there are not releases of all of these sex groupings. This meant that we could not produce species-level estimates of lifespan for each sex. Fortunately for each genus, as well as overall, we were able to produce estimates of lifespan for female-only and male-only releases (Figure 2.14).

Overall we estimate that female mosquitoes live for approximately 0.8 days longer than males (difference in posterior median estimates); female mosquitoes live for approximately
3.8 days and males for 3.0 days. An ANOVA test using the posterior median estimates from
the individual series indicated that there are no significant differences between males and
female lifespans. However, this pattern repeated across all three genera in our sample. For
*Culex* mosquitoes females lived for 0.5 days longer, on average, than males; for *Anopheles*
females lived for 1.5 days longer; and for *Aedes* the difference in lifespan was 1.6 days.
Since male mosquitoes have previously been observed to live shorter lives than females in
the lab (Styer et al., 2007) and field (Silver, 2007), we believe that our results (whilst not
statistically significant) may be indicative of this trend.

We do not know of any previous work that has modelled the effect of a mosquito lifespan
that varies by sex. However further modelling work is needed to determine the implications
of such differences in lifespan by sex to vector control efforts.

### 2.4.5 Blood- and sugar-fed mosquitoes do not live longer than those that were not fed before release

The Guerra et al., 2014 database contains data on whether the released mosquitoes were
blood- and/or sugar-fed prior to release. To ensure that we had sufficient statistical power
we estimated the lifespans of each category across all time-series in our dataset (Figure
2.15). Whilst there is a slight increase from no-feeding to sugar-only then blood-only then
sugar-and-blood (also seen in the genus-level estimates – data not shown), ANOVA tests
found that this effect was not statistically significant.

### 2.4.6 Wild mosquitoes do not appear to experience strong age-dependent mortality

We estimated six different models for mosquito mortality; each representing a previously-
used model in the literature (see ‘Methods’). Five of these models allowed for the possibility
of a mortality hazard that increases with age; the exponential model is the only model which
does not allow a variable age-dependent mortality. Here we compare the predictive perfor-
mance of each of these models using K-Fold cross-validation (see ‘Methods’). In particular
we estimate the average log probability for a single data point across our independent test
sets, with higher values for this quantity indicating a better fit of the model to the data.

There was some variability in the fit of each model across all the species (Figure 2.16).
However, in all cases the exponential model either performed statistically equivalent to, or
better than, the models that allowed for an age-dependent mortality.
Figure 2.7: Individual time-series estimates of adult mosquito mean lifespan from the MRR analysis ordered by species. The middle line in each box shows the median estimates. The left and right box whiskers show the 25%, and 75% posterior quantiles respectively. All estimates were obtained using the non-hierarchical exponential survival model as described in text.
Figure 2.8: Posterior estimates of adult mosquito mean lifespan across species, genus and overall groupings as determined from the MRR data. The middle line in each box shows the median estimates. The left and right box edges show the 25%, and 75% posterior quantiles respectively. The whiskers show the range of the data, excluding points lying more than 1.5 times the interquartile range away from each edge of the box. The numbers before the start of the left whisker indicate the number of individual time-series within each species. All estimates were obtained using the exponential survival model.
2.4. RESULTS

Figure 2.9: The estimated out-of-sample average log pointwise predictive performance of the genus- and overall-level models relative to the species-level model. Values higher up indicate a more predictive model. In all cases the predictive performance of each model was measured via K-Fold cross-validation. The upper and lower tails show the approximate 95% confidence intervals on the predictive performance relative to the species-level model.
Figure 2.10: The estimated out-of-sample average log pointwise predictive performance of the species-, genus- and overall-level models, across each species. Note the scale of the horizontal axis is -1 times the predictive performance, and hence values closest to zero indicate the best performing model. In all cases the out-of-sample predictive performance was estimated by K-Fold cross-validation.
Figure 2.11: **Posterior estimates of mean mosquito lifespan for each time-series, versus the trapping range.** The markers show the median posterior estimates, with the lower and upper bounds indicating the 25% and 75% quantiles, respectively. The black line shows a linear regression line estimated using the median posterior lifetimes, with the grey shading indicating 95% confidence intervals. All estimates were obtained using the exponential survival model.
Figure 2.12: The estimated proportion of mosquitoes surviving the EIP of malaria, chikungunya, dengue fever and Zika for the main disease vectors. Here we assumed that the EIPs were: malaria-10 days, chikungunya-2 days, dengue fever-6.5 days and Zika-12.5 days. The left and right whiskers indicate the 25% and 75% posterior quantiles, and the points indicate the mean. All estimates were obtained using the exponential survival model.
Figure 2.13: Posterior estimates of mean mosquito lifespan for each time-series versus the average monthly temperature for each study location and date. The markers show the median posterior estimates, with the lower and upper bounds indicating the 25% and 75% quantiles, respectively. The black line shows a regression line with linear and quadratic terms estimated using the median posterior lifetimes, with the grey shading indicating 95% confidence intervals. All estimates of lifespan were obtained using the hierarchical exponential survival model.
Figure 2.14: Posterior estimates of mean mosquito lifespan across female-only and male-only releases, across all studies (“overall”) and by genus. The markers show the median posterior estimates, with the lower and upper bounds indicating the 25% and 75% quantiles, respectively. The numbers before the start of the left whisker indicate the number of individual time-series within each category. All estimates were obtained using the exponential survival model. Add numbers involved in each case.
Figure 2.15: Posterior estimates of mean mosquito lifespan by pre-release feeding method across all studies. The markers show the median posterior estimates, with the lower and upper bounds indicating the 25% and 75% quantiles, respectively. The numbers before the start of the left whisker indicate the number of individual time-series within each category. All estimates were obtained using the exponential survival model. Add numbers involved in each case.
**Figure 2.16:** Average log probability for each data point from each of the estimated models of mosquito mortality. Values towards the right indicate a better predictive fit for a given model. Across all species there were no cases where the exponential model performed significantly worse than the model with the best fit. The average log probability was calculated on independent test sets using K-Fold cross-validation.
2.5 Discussion

Mosquito longevity is an important epidemiological parameter, yet one that cannot be easily inferred with MRR. The MRR experiment has for many decades been a widely used method to estimate longevity (Silver, 2007), yet the approach is often beset by low recapture rates leading to large confidence bounds on results. Furthermore, the methods used to estimate longevity from MRR data often differ, while several unavoidable assumptions may bias MRR-based results systematically. In this study we have applied common estimation procedures to data from a large number of MRR studies to reduce, as far as is possible, the sampling noise inherent in individual experiments, and to overcome the difficulty of interpreting results based on different analytical procedures. We now discuss these findings although delay comparison with dissection-based methods (Polovodova, 1949; Detinova, 1962) until Chapter 3 where we carry out a meta-analysis of such experiments. Finally, we draw on our analysis to present an appraisal of the MRR experimental procedure itself, as a means to estimating mosquito longevity.

The main results of this analysis are as follows:

1. Most mosquitoes live for less than ten days, which implies that only a small fraction of adult females can act as vectors for disease (depending on the disease and mosquito species). Although a few species live longer than this, the main vectors of malaria, dengue fever, Zika and chikungunya are all found to exhibit relatively short average lifespans (Figure 2.8).

2. The data does not reject the exponential survival model, which assumes a risk of mortality that is independent of age. However, age-dependent mortality has been detected in lab populations (Styer et al., 2007; Harrington et al., 2008; Dawes et al., 2009) and in field populations based on dissections of wild-caught adult fe-
males Clements and Paterson, 1981. This result may be due to the rarity of marked mosquitoes that live long enough to experience physiological decline, which implies that senescence may not be an important effect in the field.

3. We found no evidence of a systematic effect of temperature on mosquito lifespan (Figure 2.13). However, field observations and lab experiments have suggested the adult lifespan of *Ae. aegypti* mosquitoes is affected by temperature Jepson, Moutia, and Courtois, 1947; Yang et al., 2009; Lambrechts et al., 2011. It is unclear whether our result is because our sample has relatively few MRR experiments at extreme temperatures, or because the data lacks the statistical power to identify an effect. Nevertheless, our results suggest that the effect of temperature on adult mosquito lifespan may be relatively small in magnitude.

4. The data indicates that wild female mosquitoes live longer than males (Figure 2.14), which has been previously found in laboratory (Styer et al., 2007) and wild (Silver, 2007) populations. There is increasing interest in the biology of male mosquitoes due to the development of male-based population control methods (Ferguson et al., 2005), such as sterile insect technique (Benedict and Robinson, 2003; Alphey et al., 2010), and transgenic methods that aim to bias the sex-ratio of wild populations towards males (Lees et al., 2014; Galizi et al., 2014; Adelman and Tu, 2016). This increased focus on male mosquitoes as well as ethical concerns have likely lead to an increased frequency of MRR experiments that release only males (Figure 2.17). However we believe that experiments that elucidate aspects of the ecology of female mosquitoes continue to be of value so long as adequate measures are taken to mitigate risks to humans.

The key assumption of MRR-based methods are: (i) the behaviour and survival of
mosquitoes are not affected by their marking or handling, and (ii) the rate of mosquito
emigration from the study area is small compared to the rate of mortality. Mosquitoes
are delicate insects and it is plausible that many marking methods will impair fitness to
some degree, yet unfortunately there are seldom data to quantify this. A few studies have
measured the survival of *Ae. aegypti* and *A. gambiae* mosquitoes in laboratory conditions
after marking with a number of the fluorescent dusts commonly used in MRR studies
(Muir and Kay, 1998; Verhulst, Loonen, and Takken, 2013; Dickens and Brant, 2014). Most
experiments found that marking leads to reductions in survival (depending on the particular
dye), indicating that survival estimates based on MRR experiments will underestimate
longevity. If the assumption of limited emigration is incorrect, then the results from the
MRR analysis will also underestimate the longevity of wild mosquitoes.

Figure 2.17: The number of MRR time-series by date and sex of released mosquitoes. The numbers represent by-decade totals of the 232 time-series we used in the main analysis.
There was no correlation between estimated lifespan and trapping range (fig. 2.11), suggesting that in most MRR experiments, the study area is sufficient to preclude a major influence of emigration. However, the apparent absence of an effect of trapping range may be due to a low recapture success of outer traps in studies with a larger recapture range. Indeed we believe that our estimates are likely downwardly biased due to the effects of mosquito dispersal. This is supported by our estimates of lifespan, which suggest there are insufficient mosquitoes that survive long enough to support mosquito-borne diseases in geographies where these diseases are endemic.

Taken together, both assumptions indicate that our estimates of longevity are downwardly biased, and that our results represent lower bounds on mosquito longevity. Additional data that addresses these assumptions will clearly be very helpful in clarifying the possible extent of systematic bias in MRR based longevity estimates. Laboratory based studies are a useful approach to address the effects of marking, and we encourage entomologists to consider performing similar experiments in conjunction with future MRR experiments.

2.5.1 The statistical power of MRR studies

Consider an ideal MRR experiment, where the assumptions of no emigration and harmless marking are fully satisfied. With these data, how accurately can we estimate mosquito lifespan? To address this question we generated artificial MRR data and attempted to estimate the (known) parameters by maximum likelihood (a ‘Monte Carlo’ analysis). Specifically, we simulated releases of $N$ mosquitoes that are monitored for $m$ days in each case, and determined how statistical power to estimate lifespan depends on these two parameters (Figure 2.18). In order to focus on estimating lifespan, we assumed the mosquitoes experience constant mortality, and recaptures follow the negative binomial sampling model.
(eqn. 2.1) where the recapture probability parameters ($\psi$ and $\kappa$) were the averages of those estimated for the actual data.

![Graph](image.png)

Figure 2.18: The average error in predicting mean lifespan as a function of study length (left) and number of released mosquitoes (right) in a Monte Carlo analysis. For each parameter set we generated 400 simulated data series using the negative binomial sampling model with an exponential survival function as described in the text, and estimated the mean lifespan using maximum likelihood. The error bars show the standard deviation in the prediction error.

Unsurprisingly, the error in predicting lifespan declines as the duration of study is increased (Figure 2.18A). However once a study length is much longer than the lifespan of the majority of mosquitoes, the predictive power cannot be improved by extending the study duration. The effect of increasing release size (while holding study length constant) is similarly a case of diminishing returns (Figure 2.18B). For the parameters we use, there
are significant gains in accuracy from releasing 1,000 rather than 100 mosquitoes, but very modest gains from releasing 10,000 rather than 1000 mosquitoes.

We also conducted a power analysis for the detection of senescence in MRR experiments (Figure 2.19). Here we calculated the power of a maximum likelihood estimator of the ‘senescence parameter’ $\beta$ of the Gompertz survival function (see Table 2.3) for case study populations with three different levels of senescence (Figure 2.19A.). These indicate that the power to detect senescence is strongly dependent on study length (Figure 2.19B.), but is insensitive to release size (Figure 2.19C.). Clements and Paterson, 1981 conducted a meta-analysis of MRR and dissection-based experiments and found evidence of senescence that is, at least, qualitatively similar in magnitude to that of the ‘mild’ case considered above. For this case detecting senescence with a power of 80% requires a study length of at least 18 days. Since the median study length for experiments included in our analysis was 10 days (Table 2.1) this could partly explain our failure to detect senescence at the species level.

2.5.2 MRR design

To help inform future MRR experimental design, we investigated the influence of MRR experimental set up on recapture rate. Specifically we regressed the recapture percentage from each individual MRR time-series against a range of experimental design covariates, including both procedural factors (duration of the study, release size, collection method, trapping range etc.) and biological factors (mosquito diet prior to release, genus, sex; Table 2.5).

Surprisingly, this analysis suggests that the only reliable means of improving recapture rate is to increase the duration of a study. The number of traps was found to negatively affect the recapture rate, in contrast to our expectations. A possible explanation for this
## Table 2.5: Results from a regression analysis of determinants of recapture success rate.

Columns (1), (2), and (3) indicate different regressions. In all three cases the dependent variable was the log of the percentage of marked mosquitoes that were recaptured over the course of each MRR experiment. The standard errors for each parameter are shown in parentheses, and */** indicate significance at the 5%/1% levels. The models were estimated using ordinary least squares.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>experimental conditions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>duration of study</td>
<td>0.08**</td>
<td>0.19**</td>
<td>0.14**</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>duration of study, squared</td>
<td>-</td>
<td>-0.03**</td>
<td>-0.03**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>number released, '000s</td>
<td>-0.03**</td>
<td>-0.03**</td>
<td>-0.03**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>number of traps</td>
<td>-0.003**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>trapping range</td>
<td>-0.0003**</td>
<td>-0.0002**</td>
<td>-0.0003**</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>human landing catches</td>
<td>-</td>
<td>0.31</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.37)</td>
<td>(0.36)</td>
</tr>
<tr>
<td>CO₂/light traps</td>
<td>-</td>
<td>0.35</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.30)</td>
<td>(0.85)</td>
</tr>
<tr>
<td>aspirators</td>
<td>-</td>
<td>-0.10</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.33)</td>
<td>(0.36)</td>
</tr>
<tr>
<td>animal bait</td>
<td>-</td>
<td>0.51</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.40)</td>
<td>(0.44)</td>
</tr>
<tr>
<td><strong>mosquito ecology</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anopheles</td>
<td>-</td>
<td>-</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.33)</td>
</tr>
<tr>
<td>Aedes</td>
<td>-</td>
<td>-</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.36)</td>
</tr>
<tr>
<td>male</td>
<td>-</td>
<td>-</td>
<td>-0.76**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.27)</td>
</tr>
<tr>
<td>mix</td>
<td>-</td>
<td>-</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.31)</td>
</tr>
<tr>
<td>blood-fed</td>
<td>-</td>
<td>-</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.21)</td>
</tr>
<tr>
<td>sugar-fed</td>
<td>-</td>
<td>-</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.22)</td>
</tr>
<tr>
<td><strong>other</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-4.43**</td>
<td>-4.51**</td>
<td>-4.40**</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.47)</td>
<td>(0.53)</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td>0.23</td>
<td>0.32</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Figure 2.19: The statistical power to detect senescence for A. three case study populations as a function of B. study length and C. number released. For each parameter set we generated 500 simulated data series using the negative binomial sampling model with a gompertz hazard function ($\lambda(t) = \alpha e^{\beta t}$), and estimated the $\beta$ parameter using maximum likelihood and tested the null hypothesis $\beta = 0$ (constant mortality risk) against the alternative $\beta > 0$ using an likelihood ratio test with a 5% test size. The error bars show the standard deviation in the power for each parameter set. The recapture probability parameters ($\psi$ and $\kappa$) used were the averages of those estimated for the actual data.

result is that we are failing to adequately capture the spatial distribution of traps, by simply using the distance of the furthest trap from the release point. So as the number of traps increases, they may be distributed less densely in space, resulting in a lower recapture rate. The trapping range and release size were similarly found to negatively impact the recapture rate.

We believe that the quality of the data was not sufficient to rank the four collection methods (human landing catches, light traps, aspirators and animal bait), where based on previous observations it is expected that human landing catches likely produce the highest recapture rates (see for example Mathenge et al., 2004; Kroeckel et al., 2006). However, it is possible that differences between recapture methods are masked by not
accounting for species-level variation the efficiency of different methods are known to vary by species Mathenge et al., 2005. Male-only releases suffer from a lower recapture rate than female-only releases, which is unsurprising since most trapping methods are based on female-specific behaviour, but may also reflect their shorter lifespan.

2.6 Conclusion

In this chapter, I presented estimates of mosquito lifespan from an analysis of a database of MRR experiments. This analysis allows us, for the first time, to estimate lifespan by species and genus, and to investigate how environmental factors affect mosquito mortality. This work indicates that mosquito lifespan varies by species and genus and that female mosquitoes live longer than males. There was no evidence to suggest any systematic variation in mosquito mortality according to the average outdoor temperature of study location, however, we cannot discount the possibility that other environmental covariates, such as distance from human habitation, suitable water bodies or rainfall may affect lifespan. By fitting a range of different survival models to the data – each embodying different assumptions about mosquito mortality – this allowed us to test whether wild mosquitoes experience senescence. Across all species, we did not find any evidence that age-dependent mortality could better explain the data than the simpler exponential survival model.

In estimating mosquito lifespan from spatially-averaged MRR experiments, two assumptions are made: mosquito dispersal out of the study area is negligible compared to the rate of death; and that handling of mosquitoes and their marking does not damage them enough to affect their survival. In reality, both of these assumptions may be violated – mosquitoes have been found to migrate long distances (Dao et al., 2014) and dye has been demonstrated to significantly affect survival (Verhulst, Loonen, and Takken, 2013) – leading to estimates of lifespan that are downwardly biased. Without a gold standard
approach for estimating mosquito survival, however, it is unclear how significant these effects bias estimates of lifespan obtained from MRR experiments. Fortunately, alternative approaches do exist for determining lifespan, although admittedly, each also requires us to make assumptions about mosquito ecology. In the next chapter, I consider an approach which is based on dissection of female mosquito experiments, which allows, for some species in the dataset we collect, to compare estimates from this approach with those from MRR experiments, allowing a critical appraisal of each method.

Whilst MRR experiments provide imperfect measures of mosquito lifespan, it is worth noting that they are also often used to simultaneously determine a range of other bionomic parameters, such as dispersal distance and population densities (see Silver, 2007). Furthermore, many of these parameters cannot be determined as directly by other approaches. This means that the MRR experiment, whilst costly and variable in results, is likely to remain a mainstay approach in a field entomologist’s toolbox.
Chapter 3

A meta-analysis of mosquito dissection experiments

3.1 Abstract

An alternative method to MRR experiments for estimating lifespan of wild mosquitoes is by capturing samples of mosquitoes and dissecting the specimens. Female mosquitoes undergo predictable changes in their reproductive anatomy throughout their lives, providing a biological clock which can be used to estimate the physiological age of individual specimens when dissected. In this chapter, I describe how a dataset of female mosquito dissection experiments was compiled by a literature survey, then apply a Bayesian hierarchical model (of similar nature to that used in Chapter 2) to estimate the physiological lifespan of mosquitoes by species and genus. This analysis indicated that there were significant differences in the lifespan of mosquitoes between species and genera. To convert physiological age - a measure of the number of gonotrophic cycles a female mosquito has undergone - to calendar age, requires estimates of the length of the gonotrophic cycle. In
this chapter, I describe how a further search of the literature was undertaken to compile
a database of estimates of gonotrophic cycle duration. I then apply a statistical approach
to this data to produce estimates of gonotrophic cycle duration. These estimates are then
used to convert the physiological lifespan estimates into lifespan in terms of calendar time,
which show good correspondence with the estimates from the MRR studies. As for the
MRR analysis, I also fitted six different survival models to the data, each of which embodies different assumptions about age-dependent mortality. Unlike the MRR analysis, there
was a lack of consensus as to whether age-dependent mortality affects wild populations:
across 5 species, the exponential distribution, which assumes no increase in mortality with
age, fit the data best; in 14 others, one of the five models incorporating age-dependent
mortality improved the fit over the exponential; in the remaining 6 species, both classes
of model provided an equally good fit. Overall, I interpret these results as indicating that
the importance of age-dependent mortality varies according to a range of factors, including
climate, geography, and genetics, and is likely a subtle effect.

3.2 Introduction

Dissection-based methods to determine the physiological age of adult female mosquitoes
are the principal alternative to MRR for estimating survival and longevity (Detinova,
1962; Dye, 1992). Perhaps the simplest such method is to sample the parous rate in a
population (the proportion of adult females that have laid eggs) by separating parous from
nulliparous females (Detinova, 1962). Parous rate gives an estimate of adult female survival
by the simple formula $daily\ survival = \frac{1}{\sqrt{parous\ rate}}$ where $T$ is the duration of the first
gonotrophic cycle (Davidson, 1954). An alternative approach to determine the lifespan of
wild mosquitoes is to use changes in the insects’ morphology that occur at regular intervals
throughout life to estimate the physiological age of specimens. The most common approach
to estimate physiological age was introduced by Polovodova, 1949, and involves counting
the number of dilations in female mosquito ovarioles. In this method the maximum number
of dilations found in the ovarioles is said to be represent the parity status of an individual,
declared as the number of times a female has laid eggs. Egg-laying typically occurs once
during each gonotrophic cycle - the process of seeking a host to allow blood-feeding, followed
by digestion of the blood and maturation of eggs, and finally oviposition in a suitable site
(Beklemishev, 1940). If estimates exist for the duration of each gonotrophic cycle, then
parity status of an individual mosquito can be converted into an estimate of its calendar
age. By collecting a sufficient number of wild mosquitoes and estimating their age it is
hoped that a survival curve can be reconstructed from which we can estimate population
quantities of interest, for example, the mean lifespan.

The female reproductive system contains two paired ovaries, each of which comprises
a number of ovarioles which attach to a central tube, the calyx, which is continuous with
the oviduct (Figure 3.1) (Fox and Brust, 1994). Each ovariole consists of a growth zone
or germarium, and typically two follicles in gonoactive females. A follicle contains one
oocyte and seven nurse cells, although in follicles recently emerged from the germarium
these two cell types are phenotypically indistinguishable (Figure 3.2) (Detinova, 1962). The
developing follicle goes through a range of stages, first described by Christophers, 1911¹,
who classified these as stages I-V.

¹Sir Rickard Christophers was a giant amongst mosquito people, who worked until well after his one
hundredth birthday. As well as his significant contribution to many aspects of mosquito biology, he is
known for writing a letter to the *American Journal of Tropical Medicine and Hygiene* objecting to their
publication of his obituary (Spielman and D’Antonio, 2001).
Chapter glossary

- **Gonotrophic cycle** - the time interval between two consecutive blood meals for female mosquitoes, that encompasses (i) the search for a host and blood-feeding, (ii) the digestion of blood and egg maturation, and (iii) the search for a suitable oviposition site and oviposition (Lardeux et al., 2008).

- **Physiological age** - the number of gonotrophic cycles a female mosquito has undergone.

- **Nulliparous** - describes female mosquitoes that have yet to complete a gonotrophic cycle.

- **Parous** - describes female mosquitoes that have completed at least one gonotrophic cycle.

- **Ovariole** - One of the tubes which composes a ovary (of which mosquitoes have two) which connects to an oviduct, where immature oocytes are deposited. Ovarioles are composed of a germarium and a series of follicles.

- **I** - One oocyte in the distal portion of follicle (the end nearest the calyx) lies below seven nurse cells.

- **II** - Yolk granules appear in the protoplasm of the egg, which itself grows to become significantly bigger than the nurse cells. At the end of this stage the egg accounts for roughly half the space of the follicle.

- **III** - The share of the follicle taken up by the egg increases to roughly three-quarters the space, and the follicle elongates. The nucleus can no longer be seen through the dense yolk.
• IV - The follicle becomes longer still and the nurse cells only occupy the uppermost
tenth of the whole, with the rest of the follicle consisting of yolk.

• V - The chorion (the outermost membrane) surrounds the entire egg, and the floats
appear on each side of the egg, ready to be laid.

Whilst there still exists some controversy regarding the details of ovulation (see Section
3.5), we now describe the process as it was originally summarised by Detinova, 1962, since
this motivates the use of this method as a way to age wild female mosquitoes. During
ovulation contraction of the ovarian muscle causes the egg to move towards the distal end
of the ovariole (Figure 3.3). Here it passes through the pedicle into the lateral oviduct,
where it forms a queue with the other eggs in the common oviduct, ready to be laid.
After the eggs leave the ovariole a large empty sack remains at the point where the egg
developed. Over time the intima of the sack begins to contract, and after about a day
there remains a dilation of similar size to a follicle that has just left the gerarium. Above
the dilation sits a follicle that is ready to develop into the second egg during the next
gonotrophic cycle. The whole process is then repeated again during ovulation, with the
result that after two ovulations there now exist two distinct dilations in the ovariole. After
each subsequent ovulation another dilation in the ovarioles forms. Hence by dissecting the
ovaries and counting the number of such dilations it is possible to determine the number
of gonotrophic cycles that a mosquito has undergone.

In this chapter, we describe the collection and analysis of data from a number of exper-
iments in which wild female mosquito populations are sampled and dissected to estimate
the parity status of each individual (the number of gonotrophic cycles undergone). By
using a hierarchical Bayesian model where individual experiments from within a species
or genus are grouped together this allows estimation of mean mosquito lifespan at these
levels. These lifespan estimates are in terms of numbers of gonotrophic cycles. To convert
these into calendar ages we require estimates of the duration of each cycle. Accordingly in
Section 3.3.3 we describe the collection and analysis of data for studies that estimate the
duration of gonotrophic cycles, which is used to convert mean lifespan estimates in terms
of gonotrophic cycles to those in terms of calendar age.
Figure 3.1: The female mosquito reproductive system. Figure reproduced from University of Sydney, 2004.
Figure 3.2: The Christophers stages of ovariole development. The distal portion connects with the lateral oviduct, through which the egg passes to be laid. Figure adapted from Detinova, 1962.
Figure 3.3: The formation of dilations in ovariole development (A-J temporally). The numbers indicate the order in which the ovarioles (and the resultant dilations) develop. Figure adapted from Detinova, 1962.
3.3 Method

3.3.1 Collection of dissection data

A comprehensive search of the literature using Google Scholar (scholar.google.co.uk) was performed using various combinations of the following keywords: dissection, mosquito, parity, parous, age, and physiological age. No constraints were placed on publication date, location or type. This list was supplemented with a number of author-specific searches for those individuals most prevalent in the literature. In particular we searched for all articles authored by: Charlwood, Muller, Schlein, Samarawickrema, Reisen, Detinova, Polodova, Gillies, and Wilkes. An additional list of potential articles was provided by doing a forward article search on some of the most highly-cited articles in the database: Polovodova, 1949; Detinova, 1962; Gillies and Wilkes, 1965; Clements and Paterson, 1981. The list of results was then filtered manually by examining the titles and abstracts to produce a candidate list of the articles most likely to contain data on the physiological age of dissected mosquitoes caught in the wild, as determined by the Polovodova, 1949 method.

A relational database was used to store the raw data from the actual experiments, along with the meta-data associated with each of the experiments. In many of the published articles wild mosquitoes were caught and dissected over a period of time, and the raw data thus consists of snapshots of the age structure of the population at regular intervals in time. In the cases where the data was more sparse (fewer than ten individuals, on average, per date), we aggregated across dates and recorded this as a single entry; otherwise we recorded the snapshots of the population at each different date. We record separate series for each species that was captured, or for those that were recorded at separate capture locations (potentially with an alternative collection method), and do not aggregate over these datasets.
For each individual series we recorded the following meta-data: genus, species, collection method, whether or not insecticide was mentioned as being used in or around the time of mosquito collections, and the start and end dates of the experiment. At the article level we recorded the following meta-data: author, year, title, country, location (within country), start and end date, whether insecticide was used during any of the experimental replicates, whether a MRR experiment occurred alongside the dissection study, and the collection location (indoor or outdoor). At either the individual series or article levels we record additional meta-data describing the nature of data collection, for example explaining where the data was contained within the article, whether it was obtained by digitising graphs, and the number of separate dated series. For those few cases ($n = 16$ series) where the data was obtained by digitising graphs, we used the WebPlotDigitizer online tool (Rohatgi, 2017).

The data collection method resulted in 568 separate dissection series, across 72 published articles (see Section 7.3 for a list of the studies included in our final dataset). The published datasets cover the period from 1960-2015 with comparable numbers of studies across each decade (Figure 3.4). The statistical approach applied to the data relies on the assumption that there is a constant rate of recruitment into the adult population (see Section 3.3.2). If populations fluctuate strongly from month-to-month this assumption will likely be violated. To try to mitigate against such a risk, we aggregate the data across all dates to obtain a single series for each identifier, resulting in 201 such series. To obtain a reasonable level of accuracy on estimates we remove all those individual (aggregated) series where there are fewer than 100 mosquitoes in total in the series. Finally we remove data for any species where there was only one series resulting in 131 series across four genera ($Anopheles$, $Culex$, $Aedes$, and $Manson$) and 25 species (Table 3.1). These studies are distributed across a wide range of geographies (Figure 3.5).
### Table 3.1: The numbers of dissection series for each species or genus included in the overall dataset.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anopheles</td>
<td>gambiae s.l.</td>
<td>19</td>
</tr>
<tr>
<td>Culex</td>
<td>quinquefasciatus</td>
<td>12</td>
</tr>
<tr>
<td>Anopheles</td>
<td>maculipennis</td>
<td>11</td>
</tr>
<tr>
<td>Anopheles</td>
<td>farauti s.l.</td>
<td>10</td>
</tr>
<tr>
<td>Anopheles</td>
<td>sergentii</td>
<td>8</td>
</tr>
<tr>
<td>Aedes</td>
<td>polynesiensis</td>
<td>8</td>
</tr>
<tr>
<td>Culex</td>
<td>pipiens</td>
<td>6</td>
</tr>
<tr>
<td>Anopheles</td>
<td>culicifacies</td>
<td>5</td>
</tr>
<tr>
<td>Anopheles</td>
<td>darlingi</td>
<td>5</td>
</tr>
<tr>
<td>Anopheles</td>
<td>quadrimaculatus</td>
<td>5</td>
</tr>
<tr>
<td>Anopheles</td>
<td>stephensi</td>
<td>4</td>
</tr>
<tr>
<td>Anopheles</td>
<td>melas</td>
<td>4</td>
</tr>
<tr>
<td>Culex</td>
<td>annulirostris</td>
<td>4</td>
</tr>
<tr>
<td>Aedes</td>
<td>aegypti</td>
<td>3</td>
</tr>
<tr>
<td>Aedes</td>
<td>samoanus</td>
<td>3</td>
</tr>
<tr>
<td>Anopheles</td>
<td>minimus</td>
<td>3</td>
</tr>
<tr>
<td>Anopheles</td>
<td>rivulorum</td>
<td>3</td>
</tr>
<tr>
<td>Culex</td>
<td>thalassius</td>
<td>3</td>
</tr>
<tr>
<td>Mansonia</td>
<td>uniformis</td>
<td>3</td>
</tr>
<tr>
<td>Anopheles</td>
<td>subpictus</td>
<td>2</td>
</tr>
<tr>
<td>Aedes</td>
<td>sollicitans</td>
<td>2</td>
</tr>
<tr>
<td>Aedes</td>
<td>vexans</td>
<td>2</td>
</tr>
<tr>
<td>Anopheles</td>
<td>bellator</td>
<td>2</td>
</tr>
<tr>
<td>Anopheles</td>
<td>cruzii</td>
<td>2</td>
</tr>
<tr>
<td>Culex</td>
<td>tritaeniorhynchus</td>
<td>2</td>
</tr>
<tr>
<td>Anopheles</td>
<td></td>
<td>83</td>
</tr>
<tr>
<td>Culex</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Aedes</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>Mansonia</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>131</td>
</tr>
</tbody>
</table>
3.3. METHOD

3.3.2 Statistical analysis of dissection data

We suppose that the number of individuals recruited into the adult female population is constant over time, meaning that the age-structure of the population is stable. Whilst environmental heterogeneity will naturally lead to variation in adult recruitment over time, we hope that by aggregating data across studies undertaken over a range of dates we may reduce the impact of this effect on our results. Suppose that the probability an individual female mosquito survives until age \( a \) is given by the survival function \( S(a) \), then the number of individuals in the population surviving to this age is given by,

\[
A(a) = A(0)S(a),
\]  
\[(3.1)\]
Figure 3.5: The geographic location of each of the dissection databases included in the meta-analysis. The area of the bubbles indicates the number of unique data series available.

where \( A(0) \) is the number of female mosquitoes recruited into the population per unit time. Consider one individual experiment where we randomly sample from a wild population that is structured as per eq. (3.1). In this case the number of individuals sampled at each age is binomially-distributed,

\[
X(a) \sim \mathcal{B}(A(a), p(a)),
\]

(3.2)

where \( p(a) \) is the probability of recapturing a single mosquito of age \( a \). In what follows we typically assume that the probability of capturing a given mosquito is independent of their age, so that \( p(a) = p = \text{const} \). Since in general \( A(a) \) is large and \( p(a) \) is small, we
### Table 3.2: Summaries of the characteristics of the 131 individual dissection data series that are included in this analysis.

Note when series are censored (see Section 3.3.2) we assume that all mosquitoes dissected over the threshold are of this age when calculating summary statistics. **Add number of censored series.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Median</th>
<th>Mean</th>
<th>Max</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min age of captures, gonotrophic cycles</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>1.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Median age of captures, gonotrophic cycles</td>
<td>0</td>
<td>1.00</td>
<td>0.63</td>
<td>2.00</td>
<td>0.62</td>
</tr>
<tr>
<td>Mean age of captures, gonotrophic cycles</td>
<td>0.17</td>
<td>0.90</td>
<td>1.00</td>
<td>3.03</td>
<td>0.58</td>
</tr>
<tr>
<td>Max age of captures, gonotrophic cycles</td>
<td>1.00</td>
<td>5.00</td>
<td>5.77</td>
<td>13.00</td>
<td>2.94</td>
</tr>
<tr>
<td>Number captured</td>
<td>100</td>
<td>565</td>
<td>1,317</td>
<td>14,012</td>
<td>1,885</td>
</tr>
</tbody>
</table>

can approximate the above using a poisson distribution,

\[ X(a) \sim \text{poisson}(A(a)p). \tag{3.3} \]

However we believe that the assumption of *independent* captures of individual mosquitoes, which underlies the binomial and poisson models, is likely suspect for the same reasons as for the MRR analysis (see Section 2.3.2). As before we choose to specify a negative binomial sampling distribution that allows for non-independent captures,

\[ X(a) \sim \text{NB}(A(a)p, \kappa), \tag{3.4} \]

where we use the parameterisation such that the mean is given by \( A(a)p \), and the over-dispersion parameter, \( \kappa \), where as \( \kappa \to \infty \) the above sampling distribution approaches a poisson.

In the field, unfortunately, we do not know the number of mosquitoes recruited into the adult population, \( A(a) \), nor the probability of capturing an individual at a given point in time, \( p \). Instead we model their product \( \Psi = A(0)p \) (the population of adult female
mosquitoes of age zero that can be captured) probabilistically resulting in a model,

$$X(a) \sim NB(\Psi S(a), \kappa).$$  \hspace{1cm} (3.5)

The resultant likelihood of a data series consisting of counts: \((y(a_1), y(a_2), ..., y(a_R))\), is then calculated by assuming (conditional) independence of the observations,

$$\mathcal{L}(y(a_1), y(a_2), ..., y(a_R)|S(.), \Psi, \kappa) = \prod_{a=1}^{a_R} p(y(a)|S(a), \Psi, \kappa),$$  \hspace{1cm} (3.6)

where \(p(y(a)|S(a), \Psi, \kappa)\) corresponds to the negative binomial probability mass function

Figure 3.6: The normalised physiological age series for six species in the database. Each different coloured line represents an individual series. In each case the count for all ages has been normalised by the nulliparous (or for \(Ae. aegypti\) uniparous) count. For \(Ae. aegypti\) we start the series at uniparous individuals since there is a deficiency of nullipars for all series for this species (see Section 3.3.2). In all cases we do not include any data for censored observations (see Section 3.3.2).
for a count of $y(a)$ mosquitoes aged $a$ as specified in eqn. (3.5), and $R$ is the number of separate physiological age classes in which the count was non-zero. Since we do not know $\Psi$ we must learn it from the data. One approach to estimate this parameter could be use the number of captures of nulliparous mosquitoes, $X(0)$ in place of $\Psi$. We prefer to allow for some uncertainty in this parameter and use the data to estimate its value. However, unfortunately, the negative binomial likelihood allows too much variation in this parameter (because the data are over-dispersed), and instead we specify a likelihood of the form,

$$X(0) \sim \mathcal{N}(\Psi, \sqrt[2]{\Psi}),$$

so the first data point $X(0)$. The above allows for some freedom in $\Psi$ whilst ensuring that the parameter’s probability mass lies near enough to $X(0)$ to allow useful model estimates. $\Psi$ is set a uniform prior over the range $[0, \frac{3}{2}X(0)]$.

Since $S(a)$ is monotonically-decreasing we know that, if recruitment to the adult population is constant, then the numbers of nulliparous individuals should exceed the count in subsequent parity states. However in a number of data series there is a relative dearth of nulliparous mosquitoes versus uniparous individuals. This deficiency has been noted in a previously-published study where the authors hypothesise that it is due to the issue of sampling the nulliparous population, since they are more likely to rest outside, compared with parous individuals (Gillies and Wilkes, 1965). However there is evidence to suggest that, on average, the first gonotrophic cycle is longer than subsequent cycles (see Section 3.3.3), meaning that a relative surplus of nulliparous mosquitoes may exist in captured samples (Clements and Paterson, 1981).

In our analysis, we chose to remove the counts of nulliparous individuals from the series where the count was too low relative to uniparous or higher parity individuals since this may indicate issues with sampling the nulliparous population. Specifically we stipulated
that the nulliparous count should exceed 90% of the count for the uniparous individuals. For those cases where this condition is not met we remove the nulliparous observation and analyse the series of counts for all subsequent par (uniparous and subsequent parous states).

Here we estimate our model with one of six different survival functions (the same as for the MRR case; Table 2.3), each of which makes different assumptions regarding how the force of mortality is affected by mosquito age. To estimate lifespan at the species, genus and overall groupings we then use a hierarchical Bayesian model of the same mathematical form as for the analysis of MRR experiments (see Section 2.3.4). However the priors for the analysis of dissection data were modified so that they represented a mean prior lifespan of three gonotrophic cycles, although allowed considerable variation in mean lifespan (Figure 3.7). The priors used for the over-dispersion parameter (κ) for the negative binomial likelihood were the same as for the MRR analysis.

To determine whether mosquitoes experience age-dependent mortality we compared the predictive power of each of the models that incorporate a hazard function that depends on age with that from the exponential model. As for the MRR analysis we also used K-Fold cross-validation to perform this comparison, where the data are randomly partitioned into training and test sets (see Section 2.3.7). The model is then fitted to each training set and used to predict the data in the independent test set. Since there are fewer series than for the MRR dataset we used two partitions for each species, where each partition was of roughly the same number of individual series.

The estimates of mean mosquito lifespan that we present here (Figure 3.12) resulted from the use of the exponential survival model (i.e. no age dependence). This choice was made because we found limited evidence in support of age-dependent mortality (see Section 3.4.4). However, since the priors for each of the survival functions were specified to allow
### Survival function | Physiological age series-level priors | Group-level priors
--- | --- | ---
Exponential | $\lambda \sim \text{log-normal}(\mu_\lambda, \sigma)$ | $\mu_\lambda \sim N(-1.2, 1), \sigma \sim \text{log-normal}(-2, 1)$
Gompertz | $\alpha, \beta \sim \text{log-normal}(\mu_\alpha|\beta, 0.2)$ | $\mu_\alpha \sim N(-1.5, 1), \mu_\beta \sim N(-2.5, 0.5)$
Weibull | $\alpha, (\beta - 1) \sim \text{log-normal}(\mu_\alpha|\beta, 0.2)$ | $\mu_\alpha \sim N(-2.5, 1), \mu_\beta \sim N(-4, 0.5)$
Gompertz-Makeham | $\alpha, \beta, c \sim \text{log-normal}(\mu_\alpha|\beta|c, 0.2)$ | $\mu_\alpha \sim N(-1.4, 1), \mu_\beta \sim N(-3, 0.4), \mu_c \sim N(-4.5, 0.5)$
Logistic | $\alpha, \beta, s \sim \text{log-normal}(\mu_\alpha|\beta|s, 0.2)$ | $\mu_\alpha \sim N(-1.4, 1), \mu_\beta \sim N(-2, 1), \mu_s \sim N(-3, 1)$
Logistic-Makeham | $\alpha, \beta, s, c \sim \text{log-normal}(\mu_\alpha|\beta|s|c, 0.2)$ | $\mu_\alpha \sim N(-3, 1), \mu_\beta \sim N(-1.3, 1), \mu_s \sim N(-2, 0.5), \mu_c \sim N(-4, 0.5)$

Table 3.3: **The priors used on parameters of each different survival model for the dissection data analysis.** For the exponential model the ‘group-level’ priors were the same for the genus and ‘overall’ models that were also estimated. The exponential model was the only model that was simple enough to allow the scale parameter of the log-normal ($\sigma$) to be estimated by the data. The notation $\alpha, \beta \sim \text{log-normal}(\mu_\alpha|\beta, 0.2)$ means that $\alpha$ and $\beta$ were assigned independent log-normal priors with location parameters $\mu_\alpha$ and $\mu_\beta$ respectively, and a scale parameter of 0.2 in both cases.

for a wide variety of lifespans (Figure 3.7) the specific choice of survival function made little difference to resultant estimates (data not shown).

To demonstrate that the results we obtain are not sensitive to the particular form of hierarchical prior structure chosen we also provide estimates of the lifespan for each series analysed on its own. To do so we assume priors on the rate parameter of the exponential distribution that provides support over a wide range of possible lifespans (Figure 3.8A). We also specified a prior on the over-dispersion parameter $\kappa$ that was comparable to the hierarchical case (Figure 3.8B).

As for the MRR analysis we use Stan that implements a form of MCMC algorithm known as NUTS (Hoffman and Gelman, 2014) to sample from the posterior. The protocol followed for the MCMC sampling was the same as described in Section 2.3.6.
Analysis of censored data series

In some of the published papers the data are censored above some threshold age. In these cases we only know the number, \( y(a_T) \), of individuals who were captured and dissected with an estimated physiological age that is greater than or equal to some threshold, \( a_T \). Since we do not know the estimated physiological age of individual specimens \( i \), we represent this as a parameter, \( \mathbb{N}_i \), in our statistical model. This means that the joint distribution is a function of \( y(a_T) \) different \( \mathbb{N}_i \) parameters. Since these parameters are not directly of interest, and our chosen MCMC engine, Stan (Stan Development Team, 2014), does not directly allow discrete parameters in models, we marginalise these out of the joint distribution. To do this exactly would require an infinite sum over all the possible ages for all \( y(a_T) \) mosquitoes that have been caught whose age exceeds the threshold. Rather than carry out this intractable number of summations we instead make the approximation that all mosquitoes in this group are of the same age \( \mathbb{N}_i = \mathbb{N} \), \( \forall i \in (1, ..., y(a_T)) \). We believe this assumption is justifiable, particularly since the numbers of mosquitoes in each subsequent age category is a strongly-decreasing function of age. This means that whilst we do not know with certainty individuals’ ages, it is likely that many will be of the threshold age \( a_T \).

This approximation means that to marginalise the parameter \( \mathbb{N} \) out of the joint distribution, we are only required to do a single summation. Specifically if we have \( y(a_T) \) captured individuals of age equal to, or exceeding, some threshold age, \( a_T \), the probability of these observations is given by,

\[
q(y(a_T)|S(\cdot), \Psi, \kappa) = \sum_{\mathbb{N}=a_T}^{\infty} p(y(a_T), \mathbb{N}|S(\mathbb{N}), \Psi, \kappa) \\
= \sum_{\mathbb{N}=a_T}^{\infty} p(y(a_T)|S(\mathbb{N}), \Psi, \kappa) \times p(\mathbb{N})
\]
where \( q(y(a_T)|S(\mathcal{N}), \Psi, \kappa) \) is the probability of observing \( y(a_T) \) counts for mosquitoes of an age \( \mathcal{N} \). In practice, since we do not believe mosquitoes live for longer than 20 gonotrophic cycles (the maximum observed in the data was 13), we cut-off the summation at this point, and assume that the discrete prior probability distribution \( p(\mathcal{N}) \) is uniform over this range.

The overall likelihood for the cases where the series are censored therefore has the form,

\[
L(y(a_1), y(a_2), ..., y(a_T-1), y(a_T)|S(.), \Psi, \kappa) = \left( \prod_{a=a_1}^{a_T-1} p(y(a)|S(a), \Psi, \kappa) \right) q(y(a_T)|S(\mathcal{N}), \Psi, \kappa).
\]

(3.10)

In practice, the number of mosquitoes captured in those series that are censored represents a small percentage of total captures (the median is approximately 2%), and so the effect of the \( q(.) \) term above is likely minimal on resultant inferences.

### 3.3.3 Data collection of gonotrophic cycle duration

To convert the estimates of lifespan in physiological age into chronological age we require estimates of the duration of the gonotrophic cycle. To determine this characteristic we conducted a meta-analysis of previously-published studies that estimate the duration of the gonotrophic cycle. A search of the literature using Google Scholar (scholar.google.co.uk) was performed using the search term: ‘gonotrophic cycle duration’. The list of articles was then supplemented with a list of references discussed by Silver, 2007. Based on the abstracts of the resultant list of published studies we then decided whether to search each article for estimates of the duration of the gonotrophic cycle. Overall 79 separate estimates of this parameter were found across 42 published articles (see Section 7.4 for a list of the studies included in our dataset). Along with information about the estimates we also recorded study and series meta-data, including the location of the study, the method used for estimation, the species and genus, as well as the temperatures and/or seasons in
which the experiments were carried out. The methods used to estimate gonotrophic cycle
duration in the literature largely fell into two distinct categories: those based on MRR
studies \((n = 29)\); and those based on observations of mosquitoes in a laboratory setting
\((n = 42)\). In the MRR studies estimates are made of the duration of gonotrophic cycles
by dissecting recaptured mosquitoes at each time point using the method of Polovodova,
1949 to count ovariolar dilations. For the laboratory studies the duration of gonotrophic
cycles is determined more directly by observing the time taken for mosquitoes to mate then
blood-feed and finally oviposit.

Along with point estimates of the parameter we also collected information about the
uncertainty in the estimates (if available). In many articles the duration of the gonotrophic
cycle was estimated separately for the first versus subsequent cycles, and these estimates
were recorded separately. Raw estimates of gonotrophic cycle duration were obtained for
species across four different genera \((Aedes, Anopheles, Culex and Masonia)\), although there
was a bias towards Anopheles with \(n = 47\) estimates out of the total of 79 (Figure 3.9).

Since we collected data on the method used to estimate the duration of the gonotrophic
cycle we show the raw estimates for the two most common approaches: MRR and labora-
tory (Figure 3.10). From these results it is evident that the estimates from the laboratory
studies are, on average, higher than those using the MRR method, (although we wait until
Section 3.4.2 for a quantitative comparison).

3.3.4 Statistical analysis of gonotrophic cycle data

As discussed in Section 3.3.3, there was considerable study-level heterogeneity in the infor-
mation provided for the estimates of the gonotrophic cycle duration. A number of studies
\((n = 24)\) provided no estimates of uncertainty in gonotrophic cycle duration, whereas the
rest gave some indication of confidence or alternatively a range of possible estimates. How-
ever the types of uncertainty intervals that were specified varied considerably from study to study, from the vague (but common) ‘4-6 days’ to the more helpful ‘5±1 day (95% confidence interval)’.

The heterogeneous nature of the gonotrophic cycle estimates requires a method that explicitly accounts for this characteristic of the data. We decided to model the estimates as representing quantiles from an underlying normal distribution that represents uncertainty over possible durations of the gonotrophic cycle. In those circumstances where lower, central and upper bounds were given explicitly the data was fairly symmetric and so we believe assuming the observations come from an unskewed normal distribution is not unreasonable.

Explicitly we treated each type of uncertainty interval as follows,

- Simple range, for example, ‘4-6 days’: treat the lower and upper bounds as the 2.5% and 97.5% quantiles of a normal distribution.

- Confidence intervals, ‘5±1 day (95% confidence interval)’: treat the lower and upper bounds as the relevant quantiles of a normal distribution.

- Point estimates, for example ‘5 days’: treat this as the median of a normal distribution.

The benefit of this approach is that we can convert the quantiles from our normal distribution parameterised by a mean (μ) and standard deviation (σ) to equivalent quantiles from a standard normal,

$$ Z_{50} = \frac{G_{50} - \mu}{\sigma}, $$

(3.11)

where $Z_{50}$ and $G_{50}$ indicate 50% quantiles from a standard normal distribution and a normal distribution respectively. By manipulating the above expression we obtain the
which forms a straight line in \((Z, G)\) space with slope \(\sigma\) and y-intercept \(\mu\). Therefore by plotting the raw observations of \((Z, G)\) quantiles then estimating a linear regression line we can characterise the underlying \(\mathcal{N}(\mu, \sigma)\) distribution from which we assume they are drawn.

Due to the relative unavailability of estimates for gonotrophic cycle duration across the different genera we generated a single distribution to represent the uncertainty in this parameter by pooling all the data. We believe this is justified because ANOVA tests indicated that there was no significant variation in gonotrophic cycle duration across the different genera (Figure 3.9).

### 3.3.5 Conversion of lifespan from physiological to calendar age

The estimates of lifespan produced from analysing the dissection data are in terms of physiological age (the number of gonotrophic cycles undertaken). To allow comparison with the estimates from the MRR studies (Figure 2.8) as well as produce more useful estimates to inform disease transmission dynamical models we convert these estimates to calendar days. To do this we use the estimated parameters of the normal densities that we have assumed represent uncertainty in gonotrophic cycle duration (see Section 3.3.4) and use them to convert our posterior samples of mean physiological lifespan \((L_{p1}^1, ..., L_{pS}^S)\) to lifespan in calendar ages \((L_{c1}, ..., L_{cS})\). To do this we iterate the following for all \(i \in (1, ..., S)\) posterior samples,

1. Sample \(G_{1i} \sim \mathcal{N}(\mu_{1i}, \sigma_{1i})\), to obtain a duration for the 1st gonotrophic cycle.
2. Sample $G_{2i} \sim \mathcal{N}(\mu_{2i}, \sigma_{2i})$, to obtain a duration for subsequent gonotrophic cycles.

3. If $L_i^p > 1$, the mean lifespan is longer than one gonotrophic cycle:

   then $L_i^c = G_{1i} + G_{2i} \times (L_i^p - 1)$.

4. Else:

   then $L_i^c = G_{1i} \times L_i^p$.

In the using the above methodology to convert from physiological to calendar age we implicitly assume that for each gonotrophic cycle after the first that the cycles are of the same length (we choose one $G_{2i}$ per sample). However since we allow a different $G_{2i}$ for each sample we nonetheless believe that the above approach allows for sufficient uncertainty in the estimates of subsequent gonotrophic cycle durations. If instead we believed that significant variation in gonotrophic cycle occurred within a particular mosquito’s life (as well as between mosquitoes) then we could draw a new value of $G_{2i}$ for each subsequent cycle. This produces results with a little more uncertainty than previously (data not shown).
Figure 3.7: The priors over mean lifespan used to analyse the data for each of the hierarchical models. The distributional form of the priors is given in Table 3.3. In all cases the plots show data for 10,000 samples from the relevant prior distribution.
Figure 3.8: The priors over A. mean lifespan and B. over-dispersion parameter used to analyse the individual dissection series. The prior on the rate parameter of the exponential distribution was $\lambda \sim \text{log-normal}(-1.8, 1)$. The prior on the dispersion parameter was $\kappa \sim \text{log-normal}(1, 1)$. 
Figure 3.9: **Raw gonotrophic cycle duration estimates for A. the 1st cycle and B. subsequent cycles.** The confidence intervals represent a range of uncertainties since they were not available in a standardised form (see Section 3.3.4 for a discussion of how we handle this issue). If no statement of uncertainty was given in the estimates then we indicate this by a triangle marker. The studies are ordered according to the central estimate of gonotrophic duration.
Figure 3.10: Raw gonotrophic cycle duration estimates from A. the MRR and B. from laboratory-based studies. The confidence intervals represent heterogeneous measures since they were not available in a standardised form (see Section 3.3.4 for a discussion of how we handle this issue). If no statement of uncertainty was given in the estimates then we indicate this by a triangle marker. The studies are ordered according to the central estimate of gonotrophic duration.
3.4 Results

3.4.1 Wild mosquito lifespan in gonotrophic cycles

We first present estimates of the mean mosquito lifespan where we estimate a separate model for each data series (Figure 3.11). Whilst there is considerable variability in these estimates, in 125 of the 131 cases we estimate the median lifespan was below three gonotrophic cycles. However from this separate treatment of individual MRR time-series it is not immediately clear whether differences exist at the species or genus level. To investigate whether such group-level differences exist we estimated a hierarchical model to yield estimates at the species, genus and overall groupings (Figure 3.12). The species with the highest mean lifespan estimate was *A. sergentii*, with a posterior median of 2.5 gonotrophic cycles. This species is known as the ‘oasis vector’ or ‘desert malaria vector’ for its ability to persevere in the harsh conditions throughout the region (Sinka et al., 2010), and so it is perhaps unsurprising that both this analysis concludes that this species endures longer than most other species. At the other end of the scale is *Anopheles bellator* (median posterior estimate 0.5 cycles), which is a major vector of malaria in the Brazil’s Atlantic forests (Forattini et al., 1999; Lorenz et al., 2012). An ANOVA test using the posterior medians from each of the 131 series (from the hierarchical model) concluded that there were significant differences in mean lifespan between the species (*p* < 0.01, *F* = 3.53, *df* = 24). The result of this test remained the same if instead we used the median estimates compared to when we estimated a separate model for each series (*p* < 0.01, *F* = 2.24, *df* = 24). The Kruskal-Wallis test also concluded that this result was robust to non-normal residuals within each species (*p* < 0.01, *χ*² = 89.70, *df* = 24).

The shortest-lived genus was *Aedes*, with a mean lifespan of 0.8 gonotrophic cycles²;

---

²This was a, somewhat, surprising result, indicating that most aedines did not survive long enough to reproduce and hinting that the effective sample size may be small. This is at odds with population genetic studies which estimate a large effective population size (Lehmann et al., 1998), although these results
followed by \textit{Culex} (posterior median of 1.0 cycles); then \textit{Mansonia} (1.1 cycles); and finally \textit{Anopheles} was found to be longest with a mean lifespan of 1.4 cycles. This result was a different ordering compared with the estimates from the MRR data (Figure 2.8) which estimated that \textit{Culex} was the shortest-lived, followed by \textit{Anopheles} and \textit{Aedes}. However some of the differences in estimates can probably be attributed to the different assumptions that underpin MRR and dissection-based studies (see Section 3.5.1). An ANOVA test using the posterior medians from each of the 131 series (from the hierarchical model) concluded that there were significant differences in mean lifespan between the genera ($p < 0.01$, $F = 5.23$, $df = 3$), which was robust to the non-inclusion of \textit{Mansonia} mosquitoes, for which there were only three separate data series ($p < 0.01$, $F = 7.59$, $df = 2$). The results were again robust to the use of the data from the estimations of separate models for each of the series ($p < 0.05$, $F = 3.41$, $df = 3$). Again the Kruskal-Wallis test concluded that this result was robust to the assumption of normality for data within each species ($p < 0.01$, $\chi^2 = 25.52$, $df = 3$).

### 3.4.2 Gonotrophic cycle duration

We next present estimates of the duration of the 1st and subsequent gonotrophic cycles which were obtained by pooling data from all genera and across all experimental methods (Figure 3.13). The uncertainty in gonotrophic cycle duration was estimated as being described (by assumption) by $\mathcal{N}(4.3, 0.4)$ and $\mathcal{N}(3.8, 0.4)$ distributions for the 1st and subsequent cycles respectively. Using these distributions we estimate that the 1st gonotrophic cycle is, on average, longer by 0.5 days, although this difference is not statistically significant ($p = 0.16$).

For each of the studies we collected information on the method used to estimate may be reflective of large migratory flows between local populations rather than a high survival rate for individuals.
gonotrophic cycle duration. This allows us to estimate the duration of the gonotrophic cycle for the two predominant methods: those studies using MRR versus those using laboratory-based approaches (Figure 3.14). We determined that the uncertainty in the 1st gonotrophic cycle duration for MRR studies as a $\mathcal{N}(3.3, 0.2)$ distribution versus a $\mathcal{N}(4.8, 0.5)$ distribution for laboratory-based studies. The difference in estimated cycle length was therefore, on average, 1.5 days, and is statistically-significant ($p < 0.01$). It is not clear why the estimated gonotrophic cycle should be longer for laboratory studies, particularly since mosquitoes in the wild must locate hosts to blood-feed, and sites in which to oviposit.

### 3.4.3 Wild mosquito lifespan in calendar age

The estimates of the gonotrophic cycle duration across all species and methods (Figure 3.13) were used to convert estimates of lifespan in terms of cycles to that in terms of calendar days (Figure 3.15; see Section 3.3.5). As for the MRR analysis we find that the majority of species have a mean lifespan whose posterior median is less than 10 days (Figure 3.15). However we delay a direct comparison between the MRR and dissection-based estimates until Section 3.5.

### 3.4.4 The evidence for age-dependent mortality

We estimated models that assumed six different forms for mosquito mortality (see Table 2.3). These included the exponential model that assumes that the risk of mortality is independent of mosquito age, and other forms where we constrained the parameters in estimation so that they only allowed for a mortality that increases with age. To investigate evidence for age-dependent mortality we compare the predictive power of each of the five models that incorporated age-dependent mortality with the exponential model. As for the
MRR analysis we used K-Fold cross-validation to do this (see Section 2.3.7), and then obtained the average log pointwise predictive density for each of the models (Figure 3.16). In 14 cases we conclude that the best performing model was one that incorporated age-dependent mortality. However in the 11 other cases we either conclude that the exponential model performed best ($n = 5$) or neither the exponential nor one of the five other survival function models performed statistically better ($n = 6$). Overall we take this to mean that there is mixed evidence in favour of age-dependent mortality, particularly since we pitted the exponential model against a pool of five other models with age-dependence. And in many of the cases where the exponential model performed worse than a model with age-dependent mortality it nonetheless performed better than one or more models with age-dependence.
Figure 3.11: Individual time-series estimates of adult mosquito mean lifespan ordered by species median as determined from the dissection data. The middle line in each box shows the median estimates. The left and right box whiskers show the 25%, and 75% posterior quantiles respectively. All estimates were obtained using the non-hierarchical exponential survival model.
3.4. RESULTS

![Graph showing mosquito lifespan estimates](image)

**Figure 3.12:** Estimates of mean mosquito lifespan in terms of physiological age at the species, genus and overall levels as determined from the dissection data. The middle line in each box shows the median estimates. The left and right box edges show the 25%, and 75% posterior quantiles respectively. The whiskers show the range of the data, excluding points lying more than 1.5 times the interquartile range away from each edge of the box. The numbers before the start of the left whisker indicate the number of individual time-series within each species. All estimates were obtained using the exponential survival model.
Figure 3.13: The estimation process (A. & B.) and C. the resultant densities for gonotrophic cycle duration for the 1st (blue line) and subsequent cycles (black line). A./B. shows the quantiles of the data for the 1st/subsequent gonotrophic cycle duration versus the equivalent quantiles of a standard normal (z scores). The vertical axis intercepts and slopes from linear regression lines (orange) are then used to estimate the mean and standard deviation of the normal distribution over gonotrophic cycle duration, which are shown in C. Note that jitter in the horizontal axis has been added to plots A. and B.

Figure 3.14: The estimation process (A. & B.) and C. the resultant densities for the 1st gonotrophic cycle duration from those studies using a ‘MRR’ method (blue line) versus those with a ‘laboratory’ method (black line). A./B. shows the quantiles of the data for the gonotrophic cycle duration from A. the MRR studies and B. the laboratory studies versus the equivalent quantiles of a standard normal (z scores). The vertical axis intercepts and slopes from linear regression lines (orange) are then used to estimate the mean and standard deviation of the normal distribution over gonotrophic cycle duration, which are shown in C. Note that jitter in the horizontal axis has been added to plots A. and B.
Figure 3.15: Estimates of mean mosquito lifespan in terms of calendar age at the species, genus and overall levels as determined from the dissection data. The middle line in each box shows the median estimates. The left and right box edges show the 25%, and 75% posterior quantiles respectively. The whiskers show the range of the data, excluding points lying more than 1.5 times the interquartile range away from each edge of the box. The numbers before the start of the left whisker indicate the number of individual time-series within each species. All estimates were obtained using the exponential survival model. Samples from $N(4.3, 0.4)$ and $N(3.8, 0.4)$ distributions for the duration of the 1st and subsequent gonotrophic cycles were used to convert the estimates of physiological lifespan (Figure 3.12) into calendar lifespan.
Figure 3.16: **Average log probability for each data point from each of the estimated models of mosquito mortality for the dissection data.** We colour those species names according to whether the exponential model provided the best fit (orange) or if a model with age-dependent mortality fit best (green). Finally if neither the exponential nor any other model out-performed (i.e. if $p > 0.05$ for the pairwise comparison of average log-likelihood assuming underlying data are normally-distributed) then we coloured the name black. The average log probability was calculated on independent test sets using K-Fold cross-validation with two folds for each species.
3.5 Discussion

3.5.1 Comparison with MRR estimates

To facilitate comparison between estimates of lifespan derived from dissection-based methods and those from MRR we combine the estimates from those species and genera involved in both studies (Figures 3.15 & 2.8) into a single plot (Figure 3.17). In comparing estimates from the two approaches it is reassuring that there is considerable agreement between the estimated posteriors. In both cases we estimate that *Anopheles sergentii* was the longest lived of the anopheline species with a posterior median of 12.9 days from the MRR meta-analysis and 10.1 days from the analysis of dissection studies. This species is a major vector of malaria in the Sahara (Sinka et al., 2010), where to act as a disease vector it must persevere through these hard conditions. It is reasonable to hypothesise that this species should live longer than those in environments where the potential for blood-feeding and oviposition is greater. The species with the greatest discrepancy in the posterior estimates was the major African malaria vector *Anopheles gambiae* (Sinka et al., 2012), where we estimate a (posterior median) mean lifespan of 2.6 days from the MRR analysis and 7.7 days from the laboratory-based approaches. For the individual genera we found the greatest discrepancy for *Anopheles* (a posterior median of 3.6 days from the MRR versus 5.7 days from the dissection analysis), followed by *Culex* (2.4 versus 4.4 days). By contrast for *Aedes* the estimates from the MRR studies (posterior median of 4.5 days) exceed those of dissection-based studies (3.5 days). Across all studies we estimate from the MRR analysis that mean mosquito lifespan is 3.7 days versus 4.4 days from the dissection-based studies.

Some of the differences in these group-level estimates between the two approaches is likely due to environmental and genetic differences between mosquitoes in the experiments that were analysed in each meta-analysis. However, we believe that part of the discrepancy
can be explained by the methodological differences in approaches. In particular we specu-
late that differences in dispersal rate can explain some of the discrepancy in estimates
between the MRR and dissection-based approaches. Both Anopheles and Culex mosquitoes
are generally thought to fly farther during their lifetimes than Aedes\(^3\), meaning that the
estimates from MRR-based approaches will be most downwardly-biased for these genera.
This is supported by our results since the dissection-based estimates (themselves not reliant
on dispersal) exceed the MRR estimates for Anopheles and Culex mosquitoes, but not for
Aedes.

The key assumptions of dissection based methods are: (i) physiological age can be ac-
curately determined by dissection, (ii) the relationship between physiological and chrono-
logical age is known, and (iii) the population being sampled is in equilibrium (recruitment
matches mortality) and (iv) individual mosquitoes can be randomly sampled from the
population.

The number of gonotrophic cycles can be estimated by dissecting female mosquitoes
and counting the number of dilations in the ovarioles (Polovodova, 1949; Detinova, 1962).
However the reliability and accuracy of this method has been questioned by practitioners.
The objections include the difficulty of the dissections (Kay, 1979; Hoc and Wilkes, 1995),
the impracticality of dissecting more than a small proportion of ovarioles (Hoc and Wilkes,
1995) and the related issue of locating ovarioles that are deemed ‘diagnostic’ (Fox and
Brust, 1994), as well as the considerable variation in numbers of ovariolar dilations for
mosquitoes known to be of the same physiological age (Kay, 1979; Russell, 1986b; Hugo
et al., 2008). Indeed there is considerable uncertainty concerning the fundamental question
of how dilations in the ovarioles form. Whilst the ‘Old School’ of thought (term coined by

\(^3\)The stylised fact that aedines fly smaller distances throughout their lifetimes is, perhaps, biased towards
Ae. aegypti - a primarily urban mosquito - because others, such as Ae. albopictus originate from rural areas
(Bonizzoni et al., 2013), and have been shown to fly further in population genetic studies (Schmidt et al.,
2017).
3.5. DISCUSSION

Fox and Brust, 1994) headed by Polovodana (Polovodova, 1949) and Detinova (Detinova, 1962) considers dilations to result from normal oogenesis, a ‘New School’ headed by Lange and Hoc (Lange and Hoc, 1981) has challenged this assertion. The New School believe that only abortive oogenesis results in follicular dilations because normal oogenesis destroys the sack-like structures (Fox and Brust, 1994). This means that the success of Polovodana’s method hinges on dissecting sufficient numbers of ovarioles to discover those with the highest numbers of dilations, where abortive oogenesis has occurred in each gonotrophic cycle. They deem these ovarioles ‘diagnostic’ since only in these cases the number of dilations equals the number of gonotrophic cycles that have occurred. However as a mosquito ages the numbers of diagnostic ovarioles diminishes, since the random occurrence of normal oogenesis in a particular ovariole means it no longer has a number of dilations that equals the number of gonotrophic cycles undertaken. This increased difficulty of finding diagnostic ovarioles as a mosquito ages would mean that there is an elevated chance of age ‘hypodiagnosis’ for older specimens (Fox and Brust, 1994), and may bias estimates of population mean lifespan downwards. The difficulty of locating diagnostic ovarioles has recently been investigated using lab populations of *Culex* and *Aedes* mosquitoes by Hugo et al., 2008, who conclude that only a small percentage of ovarioles have a number of dilations equal to the number of gonotrophic cycles. Overall the bulk of evidence suggests that dissection-based physiological age determination likely understates true physiological age.

Two common methods are used to estimate the duration of gonotrophic cycles (required to convert physiological lifespan to chronological lifespan): MRR studies (see for example, Gillies and Wilkes, 1965) where marked mosquitoes are recaptured and dissected to determine the number of gonotrophic cycles that have occurred since release; and laboratory-based experiments using colonies of (typically) wild-caught females (or their progeny), whose biting and oviposition behaviour is observed over time (see for example, Afrane
et al., 2005). Whilst it is unclear how each of these methods should bias estimates of
gonotrophic cycle duration, in our analysis we find that laboratory-based studies indicate
a longer gonotrophic cycle (Figure 3.14). The distributions we ultimately used to convert
physiological age into calendar age were calculated by pooling data across both of these two
approaches, and so it is hoped our estimates incorporate this uncertainty over gonotrophic
cycle duration. Since it is unclear what biases in estimation should result from MRR or
laboratory-based studies, it is not straightforward to comment on this pooling affects our
estimates of lifespan.

If a population of mosquitoes is shrinking this leads to a relative under-abundance of
young mosquitoes, and a flattening of the survival curve, resulting in over-estimates of
lifespan, and vice versa for growing populations. For stable populations, periods when
shrinking occurs must result in equal changes in the population size compared to those
when it expands. If mosquito collections occur equally-prevalently in each of these two
modes then aggregating the data across all sampling times, and estimating a single model,
should yield an approximately unbiased estimate of lifespan. As such, to mitigate against
the risk of dynamical population structure (where possible) we aggregated series that were
gathered at different points in time. However, the additional uncertainty of a fluctuating
population size mean the credible intervals on the results are likely overly narrow.

The assumption of random sampling of the wild mosquito population has been ques-
tioned by field entomologists, although there are conflicting opinions as to whether this re-
sults in a relative paucity (Gillies and Wilkes, 1965) or abundance (Clements and Paterson,
1981) of nulliparous individuals. In our database we found a number of cases where there
was a deficit of nulliparous individuals, some of which has previously been ascribed to the
differing distribution of resting females between indoor and outdoor traps (Detinova, 1962;
Gillies and Wilkes, 1965). Indeed, differences in parity have been found for mosquitoes
collected at different times of day (Gruchet, 1962; Gillies and Wilkes, 1963; Charlwood and Wilkes, 1979), and if the mosquitoes were caught indoor or outdoors (Gruchet, 1962; Gillies and Wilkes, 1963; Gillies and Wilkes, 1965). The combined bias induced by both of these effects is unclear since individually their results are of conflicting signs. Rather than a deficit of nulliparous individuals Clements and Paterson, 1981 suppose that the longer duration of the 1st gonotrophic cycle (not supported by our analysis; Figure 3.13) should mean a relative abundance of younger individuals, and hence apply a correction factor to inflate their count of nulliparous individuals. If the population size is stationary we know that the numbers of nulliparous individuals should exceed the uniparous group. Assuming this to be the case we chose to not include those counts of nulliparous individuals in our analysis where their number was less than 90% of the uniparous. Whilst we see no obvious differences in lifespan (data not shown) according to collection method or location we cannot discount the possibility that the assumption of random sampling is violated, although the directionality of the bias induced by this is unclear. However it is likely that our uncertainty intervals for lifespan are overly narrow.

Overall the evidence suggests that the estimates of mosquito lifespan derived from dissection-based methods will likely understate true lifespan\textsuperscript{4}. This is driven mainly by the practical difficulty of dissecting sufficient numbers of ovarioles to locate those that are diagnostic of the underlying physiological state. However, the often higher estimates of lifespan from dissection-based analyses compared with those from MRR suggests that MRR estimates are more downwardly biased. Nonetheless, the different nature of the assumptions behind each of the two methods means they may offer complimentary information on mosquito survival. In addition to the differences of assumption, we note that dissection based studies require a highly specialised expertise which will often be unavailable, whereas

\textsuperscript{4}At least those using the Polovodova, 1949 method compared to the Detinova, 1962 approach, which merely determines whether a mosquito is parous or not, and requires fewer assumptions.
MRR methods are relatively simpler to use. Furthermore, most if not all dissection methods that have been used previously are only applicable to female mosquitoes, whereas MRR can be applied to either sex. Although dissection data gives detailed and, in comparison with MRR, relatively unbiased estimates of age-structure, we thus foresee a continued reliance on MRR experiments in many studies on particular mosquito populations. Efforts to use both approaches concurrently will be particularly useful, and will allow us to quantify the biases induced by the assumptions of each.

3.5.2 Comparison with laboratory-based methods

It is widely believed mosquitoes live artificially long under the comparatively benign conditions of the laboratory. We find it informative to consider estimates of lifespan derived from observations of such populations as they constitute an upper bound on the lifespan of wild populations. Also, since the numbers of mosquitoes involved in large cage experiments often numbers in the thousands, these estimates have lower uncertainty than those obtained from field experiments. Styer et al., 2007 carried out such an experiment using 45,054 female and 55,997 male *Ae. aegypti*, and found that females lived nearly twice as long as males; the median lifespan was estimated as $31.69 \pm 0.06$ days for females and $16.39 \pm 0.03$ days for males. A similar study by Dawes et al., 2009 using a lab colony of over 1000 female *Anopheles stephensi* found similar estimates for median lifespan (31-42 days). It is clear that these estimates are typically many multiples of the estimates that result from the analysis of field data. As discussed previously it is likely that both of the methods, MRR and dissection studies, which we analyse in Chapters 2 and 3, represent lower bounds on lifespan. However, without a unbiased method to measure mosquito lifespan it is difficult to quantify the gap that exists between lifespan in the field and that in the laboratory. The development of additional methods to estimate mosquito age, such as “Near-Infrared
Figure 3.17: **A comparison of the estimates of mean lifespan from the meta-analysis of MRR studies (dashed box outlines) with those from dissection-based experiments (solid box outlines).** The middle line in each box shows the median estimates. The left and right box edges show the 25%, and 75% posterior quantiles respectively. The whiskers show the range of the data, excluding points lying more than 1.5 times the interquartile range away from each edge of the box. The numbers before the start of the left whisker indicate the number of individual time-series within each species. All estimates were obtained using the exponential survival model. Samples from $N(4.3, 0.4)$ and $N(3.8, 0.4)$ distributions for the duration of the 1st and subsequent gonotrophic cycles were used to convert the estimates of physiological lifespan from the dissection-based studies (Figure 3.12) into calendar lifespan.

Spectroscopy” (NIR) which has been piloted in laboratory populations (Mayagaya et al., 2009; Sikulu et al., 2011) are likely of considerable worth here (see Chapter 4).
A better understanding of the causes of the elevated mosquito mortality hazard in the field would be beneficial in attempts to control wild populations. Presumably this is due to a range of factors that depend on the local ecology: predation, lack of food, the difficulty in finding a mate, the additional risk of mortality from blood-feeding (animals and/or humans), the relative sparsity of sites in which to oviposit, as well as variation in weather conditions. Understanding the importance of each of these factors to mosquito mortality is crucial to develop effective vector control, and we argue that field experiments to these ends are important future research directions.

3.5.3 Does age-dependence matter?

In the meta-analysis of MRR experiments we found no evidence that mosquito mortality increases with age, whereas in the dissection analysis the evidence was more mixed. We argue that this is not a fair test of the exponential model since it was pitted against a pool of five other models that incorporate age dependence. Indeed in some of the cases where age-dependent survival functions fit the data better we found that the exponential model nonetheless performed better than the worst-performing of the other models.

The absence of a finding of age-dependence for the MRR analysis could be partly explained by the shorter estimated lifespan found by this approach relative to those dependent on dissection. If marked mosquitoes live for shorter times it may be difficult to detect the relatively weak age-dependence (see below) that we find in some dissection studies. Also, it is possible that a reduction in mosquito flying ability as they age may mask the effects of age-dependent mortality as detected by MRR studies (since older mosquitoes may die at a higher rate but be more likely to remain in the study area). However, we cannot discount the possibility that the failure to detect senescence using MRR studies is due to the relatively low power of these experiments, where the experimental durations are often
3.6 Conclusion

In this chapter, I presented estimates of mosquito lifespan from a meta-analysis of mosquito dissection experiments which used the Polovodana approach of age determination (Polovodova, 1949). This method assumes that ovariolar dilations appear each time a mosquito oviposits, providing a determination of mosquito lifespan that is measured in terms of physiological...
age. Since oviposition requires a mosquito to blood feed, the results from this approach have direct epidemiological implications. If a mosquito does not blood feed at least twice in its life, then disease transmission cannot occur. This means that, rather than calendar age, mosquito physiological age, in the sense of gauging biting frequency, provides a more direct measure of disease transmission potential; particularly if mosquito gonotrophic cycle duration varies according to environmental conditions (Rúa et al., 2005; Lardeux et al., 2008). In our results, we determined that, for the majority of mosquito species, an individual female mosquito, on average, undergoes a single gonotrophic cycle during its lifetime, meaning relatively few individuals survive long enough to transmit disease. There is an ambiguity with this assertion, however. Mosquitoes that survive only a single gonotrophic cycle may still transmit disease if they are killed during or shortly after their second blood meal, before completing a full gonotrophic cycle. If the increase in mortality associated with blood feeding is substantial, then this effect may be particularly marked. Indeed, it has also been demonstrated that mosquitoes laden with Plasmodium spp. may have an elevated risk of mortality whilst blood feeding (Anderson, Knols, and Koella, 2000), since at this point in the parasite’s lifespan, there is a conflict of interests between those of the mosquito and parasite (Ewald, 1994). A mosquito would like to take a meal of sufficient size to allow egg development and fly away as rapidly as possible. Whereas, for the parasite, it is preferable for the mosquito to take a large blood meal, to provide sporozoites the greatest opportunity to be transmitted. Even if a mosquito does feed twice in its lifetime, however, it may still not be able to transmit disease. If the blood meals are taken in too rapid succession then the parasite has not the opportunity to transmute and migrate from the gut to the salivary glands (i.e. survive the extrinsic incubation period). In this sense, it is calendar age, rather than reproductive age, that matters most for disease transmission.

Calendar, rather than physiological time, is also typically the measure of time used in
mathematical or computational models of disease epidemiology. It is hoped that this work will inform future modelling work and so, we convert our estimates of mosquito lifespan in physiological age into calendar age using estimates of the duration of the gonotrophic cycle. These estimates were obtained from a meta-analysis of laboratory and MRR experiments that determined this property of mosquito populations. The results we obtained indicated that laboratory mosquitoes, on average, had longer gonotrophic cycle durations than wild individuals. This could conceivably occur if mosquitoes adjust their feeding habits according to their lifespan because, in the wild, mosquitoes survive for less time than in the laboratory. The estimates of lifespan in calendar age indicated a similar story to the results from the MRR meta-analysis. As per the MRR analysis, in the majority of species, the average lifespan was estimated as shorter than 10 days, which is generally insufficient time to transmit malaria, which has an EIP of around 10 days. Taken together, these results indicate that likely a minority of mosquitoes are responsible for the bulk burden of disease. This suggests that disease transmission should be sensitive to interventions which reduce adult lifespan.

There was a good correspondence in the estimates of lifespan obtained from each method for those species which appeared in both datasets. This similarity is encouraging as it lends support to each methodology. It is interesting, however, to comment on the differences in lifespan estimates that exist for particular cases, as this may stem from the different assumptions that underpin each approach. In Anopheles and Culex mosquitoes, the lifespan estimated from dissection experiments exceed those obtained from MRR studies, whereas for aedine mosquitoes the MRR estimates exceed those from dissections. Anopheline and culicine mosquitoes are believed to fly further in their lives than aedine mosquitoes, which may be because many of their species are generally more adapted to rural environments. Since MRR estimates assume that mosquito dispersal is negligible compared to the
effects of mortality, these differences in flight distances could explain these results. Further experimental data, however, is required to confirm this result.

Models incorporating age-dependent mortality provided a better fit to the data in most species, in contrast to the MRR experimental results. There were, however, comparable numbers of species where exponential model performed equally well or better than the collection of age-dependent models. Overall, the results from the MRR and dissection meta-analyses indicate relative weak evidence in favour of a risk of mortality that increases with age, since in most cases the exponential model fit the data as well as or better than models which incorporated senescence. This suggests that the Ross-Macdonald model, created in 1957, need not be changed to incorporate senescence, which considerably complicates the maths and likely, in reality, has only marginal repercussions on disease transmission.
Chapter 4

Using near-infrared spectroscopy to estimate mosquito population mean lifespan

4.1 Abstract

In recent laboratory experiments, it has been demonstrated that changes in cuticular hydrocarbons in mosquitoes which occur throughout their lifetime can be determined by shining near-infrared light through the mosquito specimens and examining the resultant spectrogram. The changes in mosquitoes’ body chemistry are consistent across specimens and it has been shown that machine learning methods applied to the near-infrared spectra can predict mosquito age with reasonable accuracy. In this chapter, I apply a recently developed machine learning algorithm to predict the age of mosquito specimens across a database of 13 laboratory experiments conducted on anopheline and aedine species, and demonstrate that our approach considerably bolsters individual specimen predictive per-
formance versus previously published studies. This analysis also indicates that for near-infrared spectroscopy (NIRS) to be useful for field entomological studies, considerable care to standardise experimental protocol will be necessary, since the between-study predictive performance of the machine learning algorithm is poor. Previous work has emphasised the importance of predicting the age of individual mosquito specimens, whereas in field epidemiology it is generally of more interest to estimate population mean lifespan. In this chapter, I undertake an in silico Monte Carlo experiment incorporating collection of wild mosquitoes and their subsequent analysis using NIRS, and use it to demonstrate that this technique could be used to monitor changes in mosquito lifespan that may occur during vector control campaigns.

4.2 Introduction

Mosquito borne diseases remain a major cause of suffering and death. Malaria is thought to have killed 438,000 people in 2015 (World Health Organisation, 2015) and it is estimated that the global cost of dengue exceeded $39 billion in 2011 (Selck, Adalja, and Boddie, 2014). Killing adult mosquitoes is the primary method of control for malaria, dengue, chikungunya and Zika amongst others and therefore one of the most important global public health interventions. Regular trapping of wild mosquitoes allows investigators to determine whether changes to population sizes have occurred, but mosquito abundance fluctuates substantially from day to day and according to where traps are set, rendering population size estimates imprecise. More importantly, the number of mosquitoes in itself is a poor predictor of disease transmission because many vector borne diseases have relatively long extrinsic incubation periods (see Section 2.4.2). Since many adult mosquito vectors live for less than 10 days (see Chapters 2 and 3), even in the absence of vector control it is only the rarer, older female mosquitoes that transmit the infection. Simple measures of abundance
are therefore insufficient to determine transmission risks or characterise the impact of vector control programmes. In the past, vector control was assessed by the percentage of the human population protected by interventions, such as the numbers sleeping under bednets or living in houses sprayed by insecticide. However, resistance to the most commonly-used insecticide used to control mosquito vectors of malaria is now ubiquitous throughout Africa (Ranson et al., 2011; Ranson and Lissenden, 2016), weakening the link between intervention coverage and protection.

It has long been recognised that the epidemiology of mosquito-borne disease is sensitive to wild mosquito lifespan (see Chapter 1), though there are no simple to use methods to reliably measure it in the field. Recently a number of different age-specific interventions have been proposed, including fungal biopesticides (Thomas and Read, 2007) and strains of Wolbachia that induce population mortality in older mosquitoes (Rasgon, Styer, and Scott, 2003), meaning that monitoring the age of wild populations is crucial to monitor the impact of these interventions. A key method for estimating mosquito survival in the wild is MRR studies (Chapter 2). A benefit of these experiments is that as well as estimating wild mosquito lifespan, other important ecological parameters can be estimated, such as mosquito dispersal rate or population size. However, these experiments are time-consuming, and difficult to do (Silver, 2007) and it can be costly to release the large numbers of marked mosquitoes required to achieve a reasonable level of accuracy (Desena et al., 1999). Furthermore, there are ethical concerns of releasing lab-reared or wild-caught mosquitoes into the wild due to the risk that they may contribute to the burden of disease. A more direct way of estimating mosquito age is to dissect the ovaries and assess whether (and how many times) a female has laid eggs (Chapter 3). This approach provides estimates of mosquito lifespan that are complimentary to those obtained from MRR, since the estimates do not depend on mosquito dispersal. However, this method is laborious,
meaning it is difficult to scale and doesn’t directly indicate biological age (just the number
of feeding cycles, the length of which might vary with vector control).

A number of novel laboratory methods have been developed to age mosquitoes including
the analysis of cuticular hydrocarbons (Cook et al., 2006), pteridine fluorescence (Wu and
Lehane, 1999), and near-infrared spectroscopy (NIRS) (Mayagaya et al., 2009; Liebman
et al., 2015). Unlike the other techniques, NIRS is a rapid, non-destructive method which
requires little technical training and no reagents. Large numbers of mosquitoes can be
processed relatively economically making it feasible to use NIRS to routinely assess the
efficacy of vector control interventions.

NIRS works by measuring the change in absorbance of different wavelengths of light
by organic compounds in specimens. Specifically, the stretching and bending of mainly
C-H, N-H, and O-H functional groups may cause different wavelengths of light to be ad-
sorbed according to the particular species or insect age group (Mayagaya et al., 2009).
Prior to its use in mosquitoes, NIRS has been used in a range of agricultural and pest
management applications: to identify the species present in grain stores (Dowell et al.,
1999), the sub-species of termites (Aldrich et al., 2007), to differentiate between male and
female tsetse fly pupae (Dowell et al., 2005), and to age-grade house flies (Perez-Mendoza
et al., 2002). In applying NIRS to mosquitoes, the head and thorax of specimens of known
ages are scanned (see Figure 4.1 for a picture of an near-infrared (NIR) machine) and a
simple machine learning algorithm is then used to convert spectral data (see Figure 4.2)
into estimates of biological age. Like all the above methods, NIRS has considerable mea-
surement error making predictions of individual mosquito age relatively poor. Therefore
researchers have resorted to using NIRS to define mosquitoes as young (<7 days) or old
(≥ 7 days) (Sikulu et al., 2010; Sikulu et al., 2011; Dowell, Noutcha, and Michel, 2011;
Sikulu et al., 2014; Liebman et al., 2015). Despite this simplification, NIRS is still only
able to predict the percentage of young or old mosquitoes with 78% to 90% accuracy across both male and female specimens (Sikulu et al., 2010; Sikulu et al., 2011; Dowell, Noutcha, and Michel, 2011; Sikulu et al., 2014; Liebman et al., 2015). From a practical perspective NIRS predictions of mosquito age appear to be relatively stable, with only a slight loss of precision for mosquitoes killed earlier and kept in RNAlater or other preservation methods (Sikulu et al., 2011; Ntamatungiro et al., 2013). Equally, in the laboratory mosquito age was not noticeably impacted by the physiological status of the mosquito (Ntamatungiro et al., 2013) (i.e. whether it had mated, blood fed and laid eggs) or whether or not it had been exposed to pyrethroid insecticide (Sikulu et al., 2014).

To date, a relatively simple machine learning method called Partial Least Squares (PLS; in software provided by the NIRS manufacturer) has been used to convert spectral data into
predictions of individual mosquito age. Here we collate data from several published studies that assessed the ability of NIRS to age *Anopheles gambiae* s.l. and *Ae. aegypti* mosquitoes and use it to show how a more modern method known as interval PLS (iPLS) (Norgaard et al., 2000) boosts the precision of individual mosquito age estimates. Furthermore we show that the accuracy of iPLS is consistent across all 12 datasets in our database and robust to variation in mosquito species. In the past mosquito aging techniques have been evaluated on their ability to predict individual mosquito age. However, when monitoring vector control intervention in the field, it is the overall mean and age distribution across the mosquito population that is necessary to gauge intervention efforts. To ensure unbiased estimation of population lifespan it is necessary for NIRS to produce unbiased estimates of individual age for mosquitoes of any age. In line with the results of previous studies (Sikulu et al., 2011; Dowell, Noutcha, and Michel, 2011) that used PLS we also find that
iPLS results in some biases in individual age estimates; overestimating the age of young mosquitoes and underestimating the age of old samples. Since this bias is consistent across test datasets we fit a general additive model to the errors in prediction and use this to produce unbiased estimates of mosquito age across each of our datasets. We use the bias-corrected iPLS model to demonstrate that by sampling modest numbers of mosquitoes even relatively imprecise metrics on the individual-level can generate highly accurate population-level estimates. The accuracy of predictions however is very sensitive to the nature of the calibration and testing sets. When both of these datasets were composed of mosquitoes from the same experiment, we could produce accurate predictions of mosquito lifespan. However we show that when samples from one experiment were used to calibrate our machine learning algorithm its performance on a separate test set is poor. To our knowledge this intra-experiment heterogeneity has not been discussed previously, and suggests the importance of standardising the methods for NIRS (mosquito killing, preservation, as well as instrument calibration) for field applications of NIRS.

4.3 Method

4.3.1 Data

The paired age-spectra data we analyse comes from five studies of NIRS experiments on mosquitoes produced by our experimental collaborators, Floyd Dowell and Maggy Sikulu-Lord, in experiments carried out between 2011-2015 (see Table 4.1 caption). Prior to scanning, mosquitoes were either anaesthetised and scanned fresh, or killed and preserved (see Table 1 for the preservation method used in each case). In studies A, C, D, E, F, G, H, and I, a QualitySpec Pro spectrometer (350-2500 nm; ASD Inc, Boulder, CO) was used to scan the mosquitoes to generate the spectra; in studies B, J, K and L were generated
CHAPTER 4. USING NIRS TO ESTIMATE MOSQUITO LIFESPAN

using a LabSpec 5000 NIR spectrometer (ASD Inc, Boulder, CO). In all experiments, the mosquitoes were scanned using the method described in Mayagaya et al., 2009, which we briefly summarise: the individual mosquitoes were laid on their backs 2mm underneath a 3 mm-diameter bifurcated fiber-optic probe, containing 4 collection fibers and 33 illumination fibers. The viewing area spot size was set to approximately 3mm and was focused on the thorax and head. Each individual spectrum that we used in our analysis was the result of the average of 20 repeat spectra collected by the instrument. We note that, whilst previous studies use the term ‘near-infrared’ to describe the region of the spectrum used in mosquito studies, the wavelengths considered also include parts of the spectrum in the visible light range. Our resultant database contains 4549 unique NIR-spectra measurements obtained from 13 separate experiments using lab-reared populations of Anopheles and Aedes mosquitoes of known age (Table 4.1). Whilst not knowable, the unique number of mosquitoes in the data are considerably fewer than the total suggests, since repeated observations were made from the same mosquito throughout their lifetime. The majority of these samples are for Anopheles arabiensis (n=1898), followed by Anopheles gambiae s.s. (n=1904), Anopheles gambiae s.l. (n=295), and Ae. aegypti (n=452); mostly at ages less than 15 days (see Figure 4.3).

4.3.2 Machine learning method: existent work

Previous studies have relied upon software produced by the NIRS manufacturer to build models to estimate mosquito age from individual spectra (Mayagaya et al., 2009; Sikulu et al., 2011). This software uses a statistical method known as Partial Least Squares (PLS) to estimate the age of individual mosquito samples from their respective spectra. PLS is a commonly used method in chemometrics for categorising data based on spectra. (For example, Norgaard et al., 2000 apply such a method to determine the concentration of
4.3. METHOD

### Table 4.1: A summary of the mosquito-age data for each study in our dataset.

The numbers represent the binned counts of unique mosquito-age samples within each age class. The data for experiment B are from Mayagaya et al., 2015; the data for J are from Sikulu et al., 2014; the data for K and L are from Sikulu-Lord et al., 2016; the data for experiments C, D, F, G, H, I is from Mayagaya et al., 2009; the data for experiments A and E are not yet published although the experimental protocol followed Mayagaya et al., 2009.

<table>
<thead>
<tr>
<th>Study</th>
<th>Species</th>
<th>Country of origin</th>
<th>Sex</th>
<th>Preservation method</th>
<th>Number of mosquitoes in each age group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt; 5 days</td>
</tr>
<tr>
<td>A</td>
<td><em>A. arabiensis</em></td>
<td>Tanzania</td>
<td>female</td>
<td>fresh</td>
<td>196</td>
</tr>
<tr>
<td>B</td>
<td><em>A. arabiensis</em></td>
<td>Tanzania</td>
<td>female</td>
<td>Silica gel</td>
<td>103</td>
</tr>
<tr>
<td>C</td>
<td><em>A. arabiensis</em></td>
<td>Tanzania</td>
<td>female</td>
<td>fresh</td>
<td>99</td>
</tr>
<tr>
<td>D</td>
<td><em>A. arabiensis</em></td>
<td>Tanzania</td>
<td>male</td>
<td>fresh</td>
<td>100</td>
</tr>
<tr>
<td>E</td>
<td><em>A. gambiense s.s.</em></td>
<td>Tanzania</td>
<td>female</td>
<td>fresh</td>
<td>208</td>
</tr>
<tr>
<td>F</td>
<td><em>A. gambiense s.s.</em></td>
<td>Tanzania</td>
<td>female</td>
<td>fresh</td>
<td>100</td>
</tr>
<tr>
<td>G</td>
<td><em>A. gambiense s.s.</em></td>
<td>Tanzania</td>
<td>female</td>
<td>fresh</td>
<td>100</td>
</tr>
<tr>
<td>H</td>
<td><em>A. gambiense s.s.</em></td>
<td>Tanzania</td>
<td>male</td>
<td>fresh</td>
<td>100</td>
</tr>
<tr>
<td>I</td>
<td><em>A. gambiense s.s.</em></td>
<td>Tanzania</td>
<td>male</td>
<td>fresh</td>
<td>100</td>
</tr>
<tr>
<td>J</td>
<td><em>A. gambiense s.l.</em></td>
<td>Tanzania</td>
<td>female</td>
<td>RNALater</td>
<td>91</td>
</tr>
<tr>
<td>K</td>
<td><em>An. aegypti</em></td>
<td>Australia</td>
<td>male</td>
<td>RNALater</td>
<td>46</td>
</tr>
<tr>
<td>L</td>
<td><em>An. aegypti</em></td>
<td>Australia</td>
<td>female</td>
<td>RNALater</td>
<td>41</td>
</tr>
<tr>
<td>Total</td>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td>1284</td>
</tr>
</tbody>
</table>

Figure 4.3: The number of Anopheles spectra/age samples, coloured by species. See Table 4.1 for the values used to produce this graph.
key chemical components in samples of beer!) It is a dimension reduction technique that is similar in nature to principal component analysis (PCA). PCA attempts to reduce the dimensionality of the data by decomposing the variance-covariance matrix for the independent factors (in our case, the NIR absorbance at each wavelength) into a few linear combinations of the independent variables which explain the majority of the variation. These principal components are then frequently used as features in machine learning algorithms to categorise or predict values of a dependent variable (in our case, the age of mosquitoes). In this sense, PCA is an unsupervised technique since it does not take into account the variable we seek to predict when forming the components. PLS by contrast is a supervised technique, since it explicitly takes into account the dependent variable when forming components (Maitra and Yan, 2008; Robinson, 1967). It does this by estimating latent factors in the independent and dependent variables which have the highest correlation (Figure 4.4). Whilst PLS can work with multidimensional outputs, in our case the output is a single variable (the age), although the algorithm’s mechanism remain the same. Mathematically PLS amounts to the following decomposition,

\[ X = TP' + E \]  
\[ Y = UQ' + F \]

where \( X \) and \( Y \) are matrices of the independent and dependent variables respectively (although in our case the dependent variable is univariate, and hence a vector); \( T \) and \( U \) are the projections of the independent and dependent variables into a lower-dimensional subspace; \( P \) and \( Q \) are loading matrices that give the weighting of each variable on the latent factors; \( E \) and \( F \) are error terms since the prediction of the dependent from the independent variables is, in general, not exact. A number of PLS algorithms exist to estimate the latent factors that have the highest covariance (see for example, Abdi, 2003).
Figure 4.4: A visual depiction of the method underlying Partial Least Squares. In the left-hand plot a three dimensional dependent variable is decomposed into a component $U$, which has a high correlation with the component $T$ (middle plot), that is formed from the independent variables (right hand plot). In our case the output is unidimensional, and is the age of the mosquito samples. The independent variables are measurements of NIR absorbance at each wavelength.

4.3.3 Machine learning method: our approach

We used a range of recently-developed methods to boost the predictive performance of the PLS algorithms that were used in previous mosquito age prediction experiments. The method we use is based on interval Partial Least Squares (iPLS), Norgaard et al., 2000. In this method each individual spectrum is split up into non-overlapping windows of separate wavelength intervals. The collection of individual wavelength windows are then used to train independent PLS models. The benefits of this approach are threefold: first, the individual tuning of the algorithm to each segment increases predictive performance; second, it aids understanding of the algorithms (since we are able to see how different parts of the spectra perform in predicting different characteristics of the data); lastly, this method allows us to throw away those parts of the spectra that are less useful in prediction, reducing the amount of noise in eventual predictions. After finding the most predictive parts of the spectra, it is important to look for the optimal combination of these segments. We
used the Matlab (MATLAB, 2010) Toolbox introduced by Leardi and Nørgaard, 2004 to determine the most predictive wavelength windows, and the combinations of these that had the highest accuracy. This resulted in different combinations of wavelength windows being optimal for each genus,

- Anopheles: $\lambda \in ([350 nm, 708 nm], [709 nm, 1066 nm], [1067 nm, 1424 nm])$,
- Aedes: $\lambda \in ([709 nm, 1066 nm], [1067 nm, 1424 nm])$.

We combined the models trained on the individual wavelength segments, with a PLS model trained on a wider spectrum $[350 nm, 1850 nm]$, since the inclusion of the latter was found to improve predictive performance for older mosquitoes. The predictions from the separate models were then combined using a generalised additive model with splines to result in a final estimate $\hat{\text{age}}$,

$$\hat{\text{age}} = \sum_{i=1}^{K} f_i(\hat{\text{age}}_i)$$  \hspace{1cm} (4.3)

where $\hat{\text{age}}_i$ is the estimated age from wavelength interval $i = 1, 2, \ldots, K$ and the individual $f_i()$ are functions estimated semi-parametrically, using splines. Other machine learning methods have been suggested as improvements on PLS in the literature, (in particular neural networks when the data are considerably non-linear Blanco et al., 2000) however, all other trialled methods fell short of the performance of the above methodology.

### 4.3.4 Cross validation and bias correction

We gauge the predictive performance of a model by training it on one dataset (the “training” set) and using this to predict individual mosquito ages on an independent dataset (the “test” set). The allocation of individual samples into either training or test sets was done in one of two ways: first, we generate training and testing sets by randomly selecting samples produced from a single experiment; second, we trained a model on data from one
4.3. METHOD

Figure 4.5: The partitioning of the data used to train and test our machine learning model.

experiment and used it to predict mosquito age in a test set composed of an independent experiment. The first approach allows us to gauge the within-study predictive capability of NIR, whereas the latter approach estimates between-study predictive performance.

For the within-study experiments we selected 70% of overall data to form the training set, and the other 30% comprised the test set. To avoid the risk of overfitting, we then split the training set further into a validation set (30%), and a separate smaller training set (the other 70%; see Figure 4.5). This final training set was used to train PLS models on each of the spectral windows, with the individual model predictions on the validation set used as independent variables in training the generalised additive model on the same set (and also for bias correction; see below). The test set was then used to get an independent evaluation of the performance of the final model.

To implement the iPLS model we used the Caret package for R (Kuhn, 2008). We used the default settings of the Caret package which chose a number of PLS components to minimise the RMSE across 25 bootstrapped samples of the training set. We also tried K-Fold cross-validation (using $K = 10$ folds) but the performance of the model was indistinguishable from the default bootstrapping method. For the individual wavelength interval models, the maximum allowable number of PLS components was 20 (for short
The predictive performance of the models on the test set exhibited bias for certain age ranges; overestimation for young mosquitoes and underestimation for the oldest mosquitoes (Figure 4.6A.). Since any biases in individual mosquito age predictions result in biased estimates of the population mean lifespan we sought to correct these. To do this we regressed errors in age prediction on the true age of samples in the independent validation set using a general additive model of the form,

\[
\text{error}_i = \sum_{i=1}^{K} f_i(\text{age}_i)
\]

(4.4)

where there were \( K = 5 \) smoothing components \( f_i() \), which were estimated semi-parametrically using splines. The estimated error from the above model was then used to correct the iPLS predictions on the independent test set. This process resulted in predictions that were unbiased estimates of the true age across the range of ages we surveyed (Figure 4.6B.). The cost of this bias correction was that there was a slight increase in RMSE versus the uncorrected iPLS by approximately 10%. This increase is because we trade-off a lower bias for a slightly higher variance.

The predictions of individual mosquito age that result from our machine learning approach are not restricted to be positive (Figure 4.6B.). This could be rectified by a post-prediction step where each negative mosquito age is made zero however, for our purposes this was undesirable. To produce unbiased estimates of the population lifespan we require that the estimates of mosquito age are unbiased across the range of ages encountered. For our case this means that the mean of the predicted age should be the true age across all age groups. With negative ages (whilst nonsensical biologically) this condition is met, and our estimator of mean mosquito lifespan is unbiased.
Figure 4.6: The true versus predicted age of samples using iPLS (A.) and iPLS with bias-correction (B.). The orange line is the true = predicted line and the black is a loess curve fitted to the data. The data come from study A. (Table 4.1), which used *Anopheles arabiensis*, and represent 5,821 test samples.

4.3.5 Surrogate Bayesian model

Following the results of Chapters 2 and 3 we suppose that wild mosquitoes face a constant mortality hazard (i.e. no senescence), and further assume that the population is of constant size, meaning that mosquito lifetime is exponentially distributed. We suppose we can sample randomly from this population, subject these samples to NIRS, and then use our iPLS model to predict the age of specimen. We use these NIRS-predicted ages to infer the mean mosquito lifetime of the population from which they came. Here we conduct a Monte Carlo analysis of the use of our machine learning algorithm as an estimator of this underlying wild-type population characteristic (see Figure 4.8 for a visual depiction of this process).

Since we cannot actually subject our *in silico* population of mosquitoes to NIRS we create a surrogate model that mimics the combined act of undertaking the NIRS together with prediction from our iPLS algorithm. This surrogate model takes a true mosquito age
CHAPTER 4. USING NIRS TO ESTIMATE MOSQUITO LIFESPAN

Figure 4.7: The true versus predicted age of samples for an example study along with the 95% credible interval (blue shading) for the surrogate model predictions. The orange line represents the true=predicted line. The data come from study A. (Table 4.1), which used Anopheles arabiensis, and represent 5,821 test samples.

as an input and outputs a predicted age. For the surrogate model we fit a Bayesian linear regression model with heteroscedastic errors to the combined (true, predicted) ages of the dataset resulting from NIRS predictions on the test set. Specifically we fit the following regression model to this data,

\[
predicted_i \sim \mathcal{N}(true_i, \sigma_0 + \sigma_1 true_i^2),
\]

(4.5)

where \(\sigma_0, \sigma_1 > 0\), and represented a good fit to the data (Figure 4.7).

4.3.6 Population lifespan inference

in silico populations of mosquitoes together with the aforementioned surrogate model were used to assess the performance of using NIRS to predict population life expectancy. The
4.3. **METHOD**

1. Sample individual ages from true age distribution

2. Use surrogate model to predict individual mosquito ages

3. Estimate population mean lifespan and compare with true value

---

Figure 4.8: **The Monte Carlo process used to evaluate the use of NIRS & machine learning to infer population mean age.** Across each iterate we record the difference between the true population mean lifespan and the estimated.

---

procedure for carrying out this analysis for the population mean lifetime, \( \bar{L} \) is shown in algorithm 1 (see also Figure 4.8 for further explanation of this process). Since the population age structure has an exponential age distribution, the maximum likelihood estimator of the mean lifespan is simply the sample mean age.

Here we choose \( n_{\text{iteration}} = 1000 \) which resulted in a well-characterised distribution of \( \hat{L}_i \) being obtained. In what follows we investigate how the distribution of mean lifespan estimates changes as the number of wild mosquito sampled varies.
CHAPTER 4. USING NIRS TO ESTIMATE MOSQUITO LIFESPAN

Algorithm 1 A pseudo-algorithm illustrating the method behind Monte Carlo testing of the iPLS machine learning algorithm.

1: procedure DATA GENERATION
2:   for $i = 1$ to $n_{\text{iteration}}$ do
3:     sample $(\sigma_0, \sigma_1)$ from posteriors
4:     for $j = 1$ to $n_{\text{sample}}$ do
5:       sample $true_{ij} \sim \exp(\lambda)$
6:       sample $predicted_{ij} \sim N(true_{ij}, \sigma_0 + \sigma_1true_{ij}^2)$
7:   end for
8: end for

9: procedure INFERENCE
10:   for $i = 1$ to $n_{\text{iteration}}$ do
11:     estimate $\hat{L}_i = \frac{1}{n_{\text{sample}}} \sum_{j=1}^{n_{\text{sample}}} predicted_{ij}$ (via MLE)
12: end for
13: calculate bias, variance and RMSE

4.4 Results

4.4.1 Accurate prediction of individual mosquito age within a given experiment

The machine learning algorithm (iPLS & bias-correction) accurately predicts the ages of individual samples in an independent test set which was composed of mosquitoes from the same experimental study as the training set (Figure 4.9). The predictive performance of the model was negatively correlated with the sample size, although this was not statistically significant ($\rho = -0.52$, $t = -1.9$, $p = 0.08$). We expect that the lack of statistical significance is in part related to the small sample size ($n = 12$) that we use here, since further results suggest a fairly strong dependence of machine learning performance on sample size (Section 4.4.2).

To isolate the heterogeneity in predictive performance due to idiosyncratic experimental conditions we next fixed the sample size at 200 across all experimental datasets. We then compared the performance of our machine learning algorithm with the method used by previous studies across 100 replicates (Figure 4.10). In all cases the performance of our
Figure 4.9: The true versus predicted age for each of the studies (A. to L.; Table 4.1) using test data composed of within-experiment samples. The sample size and RMSE (‘average error’) is as indicated above each panel. Studies A. to D. are were carried out with *Anopheles arabiensis* mosquitoes; E. to I. with *Anopheles gambiae s.s.*; J with *Anopheles gambiae s.l.*, and K. & L. with *Ae. aegypti*. The test data was generated by randomly partitioning the sample into training (70% of overall sample) and testing (the remaining 30%) datasets across 20 repetitions.
algorithm performed significantly better at the 5% level (Table 4.2). Overall the mean
RMSE for our machine learning method was 0.72 days less than that of the previous
method, which was consistent across species ($A.\ arabiensis$, 0.77 days; $A.\ gambiae\ s.s.$,
0.67 days; $A.\ gambiae\ s.l.$, 0.79 days; $Ae.\ aegypti$, 0.73 days).

4.4.2 Individual prediction errors decrease with sample size

To evaluate the improvement in prediction due to increased sample size we took the study
with the greatest number of samples (study A., n=871) and estimated predictive accuracy
at a range of sample sizes (Figure 4.11). As the number of samples increased the average
## 4.4. RESULTS

<table>
<thead>
<tr>
<th>Study</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15.54</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>B</td>
<td>18.52</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>C</td>
<td>10.43</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D</td>
<td>22.72</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>E</td>
<td>13.07</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>F</td>
<td>7.41</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>G</td>
<td>17.97</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>H</td>
<td>13.72</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>I</td>
<td>20.86</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>J</td>
<td>22.56</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>K</td>
<td>19.69</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>L</td>
<td>2.32</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 4.2: The results of a two sample t test comparing the mean predictive performance of our machine learning algorithm with the method used in previous studies. In all cases the alternative hypothesis was that our test performs better than the previous method. The test assumes that the data in each sample are independent, although allows for heterogeneity in variances between the samples.

Error decreased although with decreasing marginal returns. In particular, we found that the predictive error was inversely dependent on the square root of the sample size. These results suggest that there are benefits to collecting larger sample sizes, although these are most significant when increasing sample sizes from a few hundred to a thousand mosquitoes.

### 4.4.3 Inferring population mean lifespan is possible with modest sample sizes

Since our machine learning algorithm requires individual spectra to produce estimates we created a Bayesian surrogate model that took as an input the true age of a mosquito and outputted an estimate (see Section 4.3.5). This allowed us to undertake a Monte Carlo evaluation of the use of NIRS to infer population lifespan (Figure 4.8). In this analysis we assume that there is no age-dependent mortality and that there is constant recruitment into
Figure 4.11: **Increasing the number of mosquitoes used in the training dataset substantially improves the predictive accuracy of our algorithm.** The solid points represent median average error seen in Study A. (Table 4.1), and the upper and lower fences represent the 75% and 25% quantiles, each of which was calculated across 100 replicates at each sample size. The regression line here was estimated as: \( y = 0.44 + 34.81x^{-0.5} \), where \( y \) is the median RMSE and \( x \) is the number of mosquitoes in the training dataset.
the adult class, meaning that the population structure was exponentially distributed with respect to age. We assume that we can randomly sample individuals from this population and subject them to NIRS (here by inputting their true age into our surrogate model) which results in an estimated age. We also assume that alongside the wild-caught specimens we use a lab population of mosquitoes reared from wild-caught larvae as a training dataset (which hopefully mitigates against risk of poor predictive performance between experiments that we discuss in Section 4.4.5).

We assess the use of NIRS to infer mean lifespan for three different populations with mean lifespans of 3, 5 and 7 days as we vary the number of wild mosquitoes that are sampled (Figure 4.12I.). Assuming predictive accuracy of NIRS on the wild-caught mosquitoes matched that of study A, we conclude that sample sizes of 150 are sufficient to resolve between these three populations. However, it is likely that in the wild mosquitoes will possess greater heterogeneity in their physiology as they age compared to lab populations (mainly due to environmental heterogeneity). To account for this we inflate the estimated errors from study A by 100% and 200% (panels II. and III. of Figure 4.12). Even if the errors are inflated by 200% inflation NIRS can still resolve between the three populations if at least 200 wild mosquitoes are caught.

4.4.4 The bulk of uncertainty in population lifespan estimates is due to sampling variability

To assess the relative effect of sampling variation and NIRS predictive errors on the inference of the population mean we considered a hypothetical scenario of applying NIRS in the wild. In particular we assume we have a calibration dataset composed of 500 mosquitoes whose error distribution is assumed to mimic that of study A. We then suppose that we can randomly sample 100 mosquitoes from a wild population whose ages were distributed
Figure 4.12: The effect on sample size on the use of NIRS to infer mean lifespan for populations with life expectancies of 3 days (orange), 5 days (blue) and 7 days (red) assuming predictive errors from study A. (panel I.) versus inflated errors (panels II. and III.). Dashed lines indicate the true value whilst shaded area shows the 95% quantile estimates across 1000 replicates as the number of mosquitoes sampled from the population increases. In all cases we assume that the NIRS errors followed a surrogate model fit to the data from Study A. To inflate the errors we inflated \((\sigma_0, \sigma_1)\) in eqn. (4.5) by the requisite amount.
exponentially with a given mean lifespan. We then record the true age, and generate a prediction from NIRS using our surrogate model. We next computed the RMSE in prediction for the population mean lifespan from first using the true age of each sample and second using the NIRS-predicted age. We repeated this exercise across a range of population mean lifespans to obtain a decomposition of the mean lifespan prediction error into that due to sampling variation and the incremental effect of NIRS predictive error, as a function of life expectancy (Figure 4.13I.). Only for the very youngest populations (mean lifespan < 2 days) was the bulk of uncertainty in prediction due to NIRS measurement error. Whereas for the rest (mean lifespan 2+ days), the dominant source of noise was sampling variation. This is because the youngest populations have the least variation in lifespan meaning that there is little sampling variation and NIRS measurement error dominates. If instead we only sample 10 wild mosquitoes, both the sampling variation and NIRS errors become inflated, and the overall errors are disproportionately larger for populations with higher life expectancies (Figure 4.13II.).

In the wild NIRS measurement error will likely exceed that which we obtained in the laboratory. To investigate the performance of NIRS in this setting we carried out the same analysis except assuming that the overall error was 100% greater than we have found for study A. (Figure 4.13III.). In this context the relative role of sampling variation is diminished, although for populations of mean lifespans exceeding 3 days it nonetheless represented the dominant contribution to predictive uncertainty.

4.4.5 Between-study prediction is poor

The between study heterogeneity is evident from observing the individual spectra (Figure 4.14). Within a single experiment there are clear commonalities in the spectra collected, even across mosquitoes of different ages (Figure 4.14I.). However, within a single age cohort
Figure 4.13: The contribution of sampling variability (light blue shading) and NIRS measurement error (green shading) to the overall error in prediction of mean mosquito lifespan (solid dark green lines) for current NIRS errors with a sample size of (I.) 100 and (II.) 10 and (III.) a sample size of 100 with a 100% increase in errors. The data shown represent the averages across 10,000 simulations. The NIRS errors were assumed to follow those estimated for Study A.

The spectra demonstrate considerable heterogeneity (Figure 4.14II.).

To assess how experimental heterogeneity affected the predictive power of NIRS we trained a model using samples from one experiment and used this to predict ages using spectra from another. We were not exhaustive here and instead picked three indicative training and testing experimental pairs to illustrate the performance of NIRS (Figure 4.15). In all cases the average error and bias were considerably worse than the within-experiment cases (see Figure 4.9). Furthermore this poor performance did not abate when we used a training set composed of the same species as the test (Figure 4.15 left panel).

It is unclear why out of the three cases considered the performance was best when using data from study A. to predict ages in study E. Since these studies had the highest sample sizes in our dataset we investigated how sample size affects prediction for this case by varying the number of samples from study A. we include in our training set (Figure 4.16). The effect of increasing sample size was relatively modest, although it was evident that there was lower uncertainty in the predictions for each age class. The reduction in bias with increasing sample size was also not a strong effect, although for the largest
4.4. RESULTS

Figure 4.14: (I.) Example spectra for a variety of mosquito ages from within a single experiment, and (II.) example spectra for mosquitoes of age 7 days across six experiments (A. → F.). In both cases the datasets chosen represent a randomly chosen subset of the overall series.
Figure 4.15: The between-study predictive power of NIRS: using data from one study (B., A., & J.) to predict mosquito age in another (A., E., & F.). The data (blue dots) represent the results across 20 iterations, and the orange line shows the predicted = true age ideal. In all cases the entire dataset from one study was used as a training set to predict all the ages from the test experiment.

Sample size (Figure 4.16III.) there was a slight improvement in prediction for the oldest mosquitoes. Since sample size does not strongly affect our results we hypothesise that the better prediction in this case was due to a similar experimental protocol used across the two studies. Unfortunately since we do not have access to a detailed description of experimental protocols we are unable to test this hypothesis, although we hope that future work will further elucidate the impact of experimental heterogeneity on NIRS performance.
4.5. DISCUSSION

Figure 4.16: The predictive power of NIRS using data from study A. to predict ages of samples in study E. as a function of sample size (I. to III.) The data (blue dots) represent the results across 20 iterations, and the orange line shows the predicted = true age ideal. In cases I. and II. in each iteration a random subset of the data from study A. was chosen as training data.

4.5 Discussion

NIRS can generate accurate estimates of the average age of mosquito populations despite predictions of individual mosquitoes being relatively uncertain. This makes NIRS a promising tool for directly monitoring the efficacy of vector control interventions that targets adult mosquitoes. The ability of an insect vector to transmit disease (the vectorial capacity) is highly sensitive to the death rate of the mosquito (Macdonald, 1957; Smith and McKenzie, 2004; Brady et al., 2016). This makes average mosquito age an important metric for assessing community-level protection and the likelihood of a disease outbreak.

Enthusiasm for the use of NIRS (and other aging methods) has been tempered by the relatively high measurement error seen in published studies. All previous work has used laboratory reared mosquitoes which are likely to be more uniform (and give more consistent spectra) than those caught in the wild due to differences in diet, environment and diversity of hazards. It is therefore assumed that NIRS precision will further diminish once these real-life factors are taken into consideration. Here we show that NIRS can generate unbiased estimates of mosquito age which means that it can generate accurate
population-level estimates of average mosquito life-expectancy. Shifting the scale at which mosquito age-grading methods are evaluated from the individual-level to the population makes epidemiological sense as this is likely to be most important for most uses of the technology. This work therefore shows that NIRS is very well suited to monitor wild mosquito populations. More accurate methods of aging individual mosquitoes (with lower measurement error) will make relative little difference to the precision of population-level estimates. This is especially the case in mosquito populations with an average life-expectancy of 3-7 days, which is the range of age distributions mathematical models predict are going to be the most informative for the control of malaria (White et al., 2011) and likely for other mosquito-borne diseases also.

New methods for aging vectors should therefore concentrate on minimising bias to ensure the technology can be used at a population-level. It is also imperative to evaluate the NIRS in more natural field settings before its use as a tool for routinely monitoring vector populations can be assessed. It is reassuring that simulations which artificially increase the individual error of mosquito age estimates did not substantially reduce overall mean population precision. The average age of wild caught mosquitoes will also be sensitive to biases in trapping techniques, as some methods are thought to preferentially catch certain aged mosquitoes (see for example, Gillies and Wilkes, 1965). These biases need to be quantified and adjusted for in any mean estimate of population age. This work also assumed that mosquitoes die at a constant rate in the wild and the models would need to be adjusted to account for senescence should evidence arrive that they senesce.

Future studies using NIRS should carefully consider the structure of the calibration dataset to improve predictive accuracy. Increasing the number of mosquitoes improves precision though care should be taken to ensure a broad range of ages, each with the same number of mosquitoes. In the studies analysed here each had fewer older mosquitoes.
It is therefore not surprising that the naïve model better fitted young mosquitoes and underestimates the age of older vectors. Given the epidemiological importance of older mosquitoes a more balanced calibration dataset would be advisable.

Further work is also needed to improve the machine learning methods used to convert spectral data into mosquito age. There are a multitude of different predictive learning methods currently being developed, many of which may be better suited to analysing these data than the relatively simple techniques employed here. For example, it is possible that using deep learning methods (see for example, LeCun, Bengio, and Hinton, 2015) could generate significant improvements to individual predictions.

This study is the first-time data from different laboratory reared mosquito populations are analysed at the same time. Unfortunately it is clear that a calibration dataset generated with one population of mosquitoes cannot be used to predict the age of a different mosquito population. Further work is needed to test whether this is true as the data analysed here were from a disparate selection of studies which span multiple years, locations, different spectrograms, mosquito killing techniques and different mosquito sub-species. Further work that standardises the procedures or refining the machine learning methods may enable this intra-study variability to be reduced, possibly allowing a universal calibration dataset. If not then every study may require their own group of mosquitoes with known age in order to train the machine learning algorithm. These calibration datasets could be generated in the field using the F1 generation of wild caught mosquitoes which would encompass the genetic heterogeneity seen in the wild. Nevertheless the increased investment required to do this would diminish the utility of NIRS for routine vector monitoring, though it would be still invaluable as a research tool.
4.6 Conclusion

In this chapter, I applied recently developed machine learning methods to predict the age of individual mosquitoes in near-infrared spectroscopy experiments from a database of previously published studies, which was compiled by my collaborators and me. This statistical approach allowed a significant improvement in the predictive error for individual specimens over previously published methods. Unfortunately, however, whilst the NIR spectra were predictive of mosquito ages within a given experiment, models trained on data from one experiment could not be used to predict the ages of mosquitoes in another. The source of poor between-experimental predictive power is not currently knowable since the experiments used differing spectroscopy machines, preparation methods and were conducted on different laboratory strains of mosquitoes. If the experimental methodology drives the poor predictive power then, standardising experimental protocols would make NIRS an attractive method for entomological monitoring. By contrast, if genetic differences in the mosquitoes or environmental covariates are the source of low predictive power between experiments, then NIRS would require a separate calibration set for each application. Whilst this may be possible in some circumstances, this would diminish the field applicability of NIRS. Further experiments, therefore, to determine the source of between experiment heterogeneity in NIR spectra will be crucial.

Epidemiologists are primarily concerned with estimates of population lifespan. To prioritise minimising bias in this quantity we modified the machine learning method to incorporate a bias-correction step. Mean estimates can then be improved by increasing the number of mosquitoes sampled. This trade-off between precision and bias therefore depends on the ease with which mosquitoes can be caught in the wild. This work indicates that the number of mosquitoes required to generate useful estimates are well within the range that can be feasibly collected by routine vector monitoring programmes (see Tables
NIRS has already proved its ability to differentiate between morphologically indistinguishable *An. gambiae s.s.* from *An. arabiensis* in wild caught mosquitoes from Tanzania (Sikulu et al., 2010) with relatively good accuracy. Further work is needed to test other vector species in different locations and to determine whether its predictive accuracy can be improved by refining the machine learning methods. At the moment PCR is expensive and require specialized training, restricting their use for routine monitoring. If so then NIRS would be ideally placed to be a single method capable of estimating mosquito age, species and infection status in mosquito populations.
Chapter 5

Using a homing endonuclease to control vector populations in a landscape with temporal heterogeneity

5.1 Abstract

The use of gene drive systems to manipulate populations of malaria vectors is currently being investigated as a method of malaria control. One potential system uses driving endonuclease genes (DEGs) to spread genes that impose a genetic load. Previously, models have shown that the introduction of DEG-bearing mosquitoes could suppress or even extinguish vector populations in spatially-heterogeneous environments which were constant over time. In this chapter, I combine a stochastic spatially-explicit model of mosquito ecology with a rainfall model which enables the generation of a variety of daily precipitation
patterns. I then use the model to investigate how releases of a DEG that cause a bias in population sex ratios towards males are affected by seasonal or random rainfall patterns. The parameters of the rainfall model are then fitted using data from Bamako, Mali, and Mbita, Kenya, to evaluate release strategies in similar climatic conditions.

5.2 Introduction

Vector control lies at the forefront of efforts to combat malaria, and is expected to have continued importance for disease control in the foreseeable future (World Health Organisation, 2015; World Health Organisation, 2016a; World Health Organisation, 2017a). Insecticides, delivered via bednets and indoor residual spraying, are the current mainstay of vector control and their widespread application has helped bring about dramatic reductions in malaria incidence over recent years (World Health Organisation, 2015). Insecticide-based programmes alone, however, are not expected to be sufficient to eliminate malaria from the worst affected parts of sub-Saharan Africa, where climatic and socio-economic conditions are particularly conducive to endemic transmission (Le Menach et al., 2007; Ferguson et al., 2010b). Moreover, the emergence of insecticide resistance in vector populations threatens the continuing efficacy of these measures (Read, Lynch, and Thomas, 2009; Ranson et al., 2011; Ranson and Lissenden, 2016). Thus there is considerable need for innovative complimentary methods of vector control (Tanner et al., 2015).

Homing endonuclease genes (HEGs) (and more generally, other driving genetic constructs - DEGs - such as those involving Crispr-Cas9; Godfray, North, and Burt, 2017) offer potential for vector control by enabling the genetic manipulation of vector populations (Burt, 2003). HEGs spread rapidly in populations by using a process known as homing, where heterozygous cells are converted to homozygotes (Mueller et al., 1993). They are found in many single-celled organisms, and are typically associated with self-
splicing introns or inteins, that catalyse their own removal from mRNA, tRNA and rRNA, meaning that there is low toxicity to the host organism (Chevalier and Stoddard, 2001). For use in vector control, a HEG could be inserted into a functional mosquito gene in order to induce a population-wide knockout of that gene (the so called classic-HEG approach). Alternatively, a HEG could be inserted onto the male-specific Y-chromosome so that it disrupts the female-specific X-chromosome and thus biases the sex-ratio towards male offspring (Y-drive). Both approaches aim to either suppress or eliminate the target vector population.

The potential of these approaches has been demonstrated in laboratory conditions (Windbichler et al., 2011; Galizi et al., 2014), and investigated using mathematical models (Burt, 2003; Derecdec, Burt, and Godfray, 2008; Derecdec, Godfray, and Burt, 2011; North, Burt, and Godfray, 2013). The earlier models considered HEG releases in large panmictic populations to help understand the population load a HEG could impose in idealised conditions (Burt, 2003), how this depends on the specific nature of the HEG deployed (Derecdec, Burt, and Godfray, 2008), and the implications for malaria reduction (Derecdec, Godfray, and Burt, 2011). More recently, North, Burt, and Godfray, 2013 investigated how the spread of a HEG may be affected by the spatial structure of natural populations in heterogeneous landscapes. This work suggests that spatial variation in habitat quality does not affect the likelihood of spread in landscapes with abundant mosquito resources. However, in sparse landscapes a HEG may become extinct before reaching all parts of the population, meriting the use of multiple release site.

These models provide a foundation for exploring how HEGs might influence vector populations, yet have not considered the role of temporal variance in rainfall. This may be important because members of the *Anopheles gambiae* complex, which include many of the major malaria vectors in Africa, tend to develop as larvae in either semi-permanent or
ephemeral water bodies created by rain (Gimnig et al., 2001; Shililu et al., 2003; Fillinger et al., 2004). The population dynamics of these species are therefore, to varying degrees, sensitive to fluctuations in rainfall (Shililu et al., 2003; Fillinger et al., 2004; Koenraadt, Githeko, and Takken, 2004; Fillinger and Lindsay, 2011).

The aim of this paper is to investigate how fluctuations in rainfall will influence the performance of a Y-drive HEG. We introduce a flexible model of daily precipitation which enables the generation of artificial rainfall data with specified characteristics (Bárdossy and Plate, 1992). This is combined with the mosquito population model of North, Burt, and Godfray, 2013 to explore HEG releases in different climates. In particular, we address the roles of seasonality and random fluctuations in rainfall and how these forms of variability should be accounted for when planning releases. We use two case-studies, based on Bamako in Mali and Mbita in Kenya, to investigate the combined influence of seasonal and random variability in these locations. These locations were chosen because they exemplify two distinct climatic regions of Africa where malaria incidence is high and HEGs may eventually be deployed (the Sahel and tropical east Africa respectively), and in each case we use historical data to fit our rainfall model. In summary, our analysis addresses the following questions. 1. When is the best time to release a HEG in a strongly seasonal environment? 2. How will shorter-term random variability in rainfall affect the performance of a HEG? 3. What is the optimal release strategy in the two case-study locations?

5.3 Method

5.3.1 Mosquito population model

Our model extends the spatially explicit model of Anopheles population dynamics developed by North, Burt, and Godfray, 2013, by incorporating temporal variability in environ-
mental conditions. Since the demographic component of both the current model and its predecessor are the same, we give only a brief review of the assumptions here. The model considers a population of mosquitoes that are each characterised by their sex, life-stage (juvenile or adult), genotype (whether or not they possess HEGs), and location in two-dimensional space. The landscape consists of small water bodies (aquatic habitats), and houses (feeding sites). The simulation is continuous time, where time steps are decided by the Gillespie algorithm (Gillespie, 1977). The values of demographic and environmental parameters used in the simulations are as detailed in Table 5.1.

Juvenile mortality can occur at any point within the simulation, and is increasing in the number of larvae at a particular aquatic habitat. We assume that density-dependent competition for food only acts on the instantaneous juvenile mortality rate, but does not affect the developmental rate, nor adult mortality. Whilst the latter two effects have been observed (Gimnig et al., 2002; White et al., 2011; Muriu et al., 2013), for simplicity we do not model these effects, although this could be considered in future work. The rate of mortality from juvenile competition is given by $\alpha n_{J_i}(t)$, where $n_{J_i}(t)$ is the number of juveniles at aquatic habitat $i$ at time $t$. We also assume that the length of the juvenile stage is fixed at 9 days. Further work might relax this assumption by considering the individual stages of juvenile development: eggs, larvae (within this stage into instars) and pupae (Hancock and Godfray, 2007).

Emergent adult females undergo gonotrophic cycles, and are categorised according to whether they are searching for feeding sites in order to blood-feed, or aquatic habitats, in which to oviposit. Adult female mosquitoes move through the landscape at rates governed by their proximity to objects of search, representing area-restricted search behaviour. Dispersal is controlled by three parameters: a jumping rate $r$, the strength of the reduction in movement near an object of search, and the maximum jump distance $s_G$. A host-seeking
female jumps at rate $r/\beta C_H(t)$, where $C_H(t)$ is the number of houses within the detection distance $s_H$ (when there are no houses within a distance, the jumping rate is $r$). When a jump is made, a new location is drawn uniformly at random within a radius $s_G$ centred on the mosquito’s current location. Oviposition occurs in the same manner, except that a count is made of the number of aquatic habitats $C_O(t)$ in the local vicinity of the mosquito.

Mating occurs only between a male and unmated female which are within a given distance $s_M$ of each other. The mating rate of a given female is given by $m C_M(t)$, where $C_M(t)$ is the number of males within radius $s_M$ of her current location at time $t$. Blood feeding occurs at a rate $\gamma_H C_H(t)$ and, afterwards, the individual female becomes endowed with $q$ eggs, where $q$ is drawn from a Poisson distribution with mean $\kappa$. Oviposition occurs at a rate $\nu C_O(t)$, and during an oviposition, a female deposits $p_X$ female and $p_Y$ male eggs into a given aquatic habitat, where $p_X \sim \text{Poisson}(\omega/2)$ and $p_Y \sim \text{Poisson}(\omega/2)$, if the female carries wild-type sperm and $p_X \sim \text{Poisson}(\omega/2(1-e))$ and $p_Y \sim \text{Poisson}(\omega/2(1+e))$, if the female carries transgenic sperm, where $e$ is the cleavage rate.

Following North, Burt, and Godfray, 2013, we assume that adult males do not disperse, and are confined to areas near aquatic habitats. This assumption was made because the behaviour of males, for example, where swarms form, and how individuals move between them, is not well understood. We are currently, however, working towards a model where both sexes of mosquito move. The males are categorised by whether their Y-chromosome carries the HEG, and mated females by whether the sperm they carry is X, Y or Y-HEG. The HEG biases the offspring towards males, by shredding 85% of the X-chromosomes in randomly-selected gametes, reducing the proportion of viable females.

The landscapes, which consisted of toroidal two-dimensional space, were populated at the start of each simulation with a stochastically-generated distribution of aquatic habitats and houses. The houses occurred at static locations in the domain, whereas the aquatic
5.3. METHOD

habitats were created dynamically according to a spatial Poisson process (see Section 5.3.2). North, Burt, and Godfray, 2013 investigated the effect of spatial covariance between the location of aquatic habitats and houses, finding that the effects of low aquatic habitat density are amplified or mitigated if this covariance is negative or positive, respectively. Here, since we are most interested in the effects of temporal covariance, we assume that the spatial covariance between aquatic habitats and houses is zero.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Default value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egg laying rate</td>
<td>ν</td>
<td>0.12</td>
<td>$day^{-1} C_{O}^{-1}$*</td>
</tr>
<tr>
<td>Feeding rate</td>
<td>γH</td>
<td>0.015</td>
<td>$day^{-1} C_{H}^{-1}$**</td>
</tr>
<tr>
<td>Egg load after feeding</td>
<td>κ</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Rate of mortality from competition</td>
<td>α</td>
<td>0.05</td>
<td>$day^{-1} n_j^{-1}$</td>
</tr>
<tr>
<td>Mortality rate in adult stages</td>
<td>μ</td>
<td>0.1</td>
<td>$day^{-1}$</td>
</tr>
<tr>
<td>Mating rate</td>
<td>m</td>
<td>0.01</td>
<td>$day^{-1} C_{M}^{-1}$***</td>
</tr>
<tr>
<td>Rate of HEG cleavage</td>
<td>e</td>
<td>0.6</td>
<td>generation$^{-1}$</td>
</tr>
<tr>
<td>Mean number of eggs laid per oviposition</td>
<td>ω</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Maximum mating distance</td>
<td>$s_M$</td>
<td>0.3</td>
<td>km</td>
</tr>
<tr>
<td>Feeding site detection distance</td>
<td>$s_H$</td>
<td>0.5</td>
<td>km</td>
</tr>
<tr>
<td>Breeding site detection distance</td>
<td>$s_O$</td>
<td>0.15</td>
<td>km</td>
</tr>
<tr>
<td>Basic jump rate</td>
<td>r</td>
<td>10</td>
<td>$day^{-1}$</td>
</tr>
<tr>
<td>Maximum jump distance</td>
<td></td>
<td>0.3</td>
<td>km</td>
</tr>
<tr>
<td>Stepping reduction near search object</td>
<td></td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Default value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation domain area</td>
<td></td>
<td>16</td>
<td>km$^{-2}$</td>
</tr>
<tr>
<td>Mean density of feeding sites</td>
<td></td>
<td>21.5</td>
<td>km$^{-2}$</td>
</tr>
<tr>
<td>Mean density of aquatic habitats</td>
<td>κ</td>
<td>varies between 100-240</td>
<td>km$^{-2}$</td>
</tr>
<tr>
<td>Breeding site turnover rate</td>
<td>χ</td>
<td>0.1</td>
<td>$day^{-1} site^{-1}$</td>
</tr>
<tr>
<td>Mean density of sample points</td>
<td></td>
<td>100</td>
<td>km$^{-2}$</td>
</tr>
<tr>
<td>Sample point turnover rate</td>
<td></td>
<td>0.02</td>
<td>$day^{-1}$</td>
</tr>
</tbody>
</table>

Table 5.1: The parameter values used in the rainfall simulations. The local densities $C_O(x,y)$, $C_H(x,y)$ and $C_M(x,y)$ at position $(x,y)$ are the number of aquatic habitats, feeding sites and mates within distances $s_O$, $s_H$ and $s_M$ of that point respectively. The references for using these parameter values are given in North, Burt, and Godfray, 2013.
CHAPTER 5. USING HEGS TO CONTROL VECTOR POPULATIONS

5.3.2 Linking rainfall with aquatic habitat density

The model was previously used to explore how landscape structure will influence the use of HEGs, by simulating releases across a wide range of feeding and aquatic habitat distributions (North, Burt, and Godfray, 2013). The analysis considered only static feeding site distributions, yet the aquatic habitats were allowed to dynamically appear and disappear to reflect the observation that *Anopheles* larvae often develop in small water bodies subject to rainfall and evaporation (Gimnig et al., 2001; Shililu et al., 2003; Fillinger et al., 2004). North, Burt, and Godfray, 2013 assumed that the rates of site creation and destruction were unchanging and equal in any given landscape, so that the average number of sites per unit area remained constant through time.

To investigate the role of climate, we suppose that variation in rainfall influences the creation of aquatic habitats. We denote by $\theta_d$ the rainfall amount on day $d$ and we will refer to $\{\theta_d\}$, the time-series of daily rainfall, as the weather. The population model is linked to weather by supposing that on a given day $d$, aquatic habitats are randomly created at a rate $\chi \times \kappa \times \theta_d$ per unit area, which means that the number of sites created on day $d$ is poisson distributed with expectation also equal to $\chi \times \kappa \times \theta_d$ per unit area. Sites are destroyed at rate $\chi$ per site, which means the longevity of a given site is exponentially distributed with mean $\frac{1}{\chi}$. Note that a large value of $\chi$ means aquatic habitats appear quickly during rain and disappear quickly during drought, and we thus use this parameter to control the lag between rainfall and aquatic habitat density. The parameter $\kappa$ will be used to control the long-run average density of aquatic habitats, allowing us to investigate HEG releases across landscapes that share similar climates yet differ in their overall extent of breeding habitat.
5.3. METHOD

5.3.3 A general model for rainfall

We model the rainfall dynamics using the procedure first described by Bárðossy and Plate, 1992. We consider a time-series \( \{W_d\} \) that represents the net sum of meteorological processes influencing rainfall that we will (subsequently) transform to yield the rainfall time-series \( \{\theta_d\} \). Rather than attempting to model the meteorological processes explicitly, we assume that on each day \( d \), \( W_d \) is a random variable that depends on three factors, (i) the conditions of the previous day \( (W_{d-1}) \), (ii) the trend in meteorological conditions \( \{\mu_d\} \), and (iii) random noise \( (\epsilon_d \sim \mathcal{N}(0, \sigma)) \). Specifically, \( \{W_d\} \) is a first-order autoregressive (AR-1) process given by the equation,

\[
W_d = \rho W_{d-1} + (1 - \rho) \mu_d + \epsilon_d,
\]

(5.1)

where the parameter \( \rho \) controls the degree to which conditions are dictated by recent history rather than the underlying trend \( \{\mu_d\} \). We will use the standard deviation of the noise term \( \sigma \) to control the extent of randomness in conditions. Since the time-series \( \{W_d\} \) may shift between positive and negative values, it is necessary to transform the series to derive the rainfall time-series \( \{\theta_d\} \). Following Bárðossy and Plate, 1992, we use the transformation,

\[
\theta_d = \begin{cases} 
W_d^\beta, & \text{if } W_d \geq 0 \\
0, & \text{otherwise}
\end{cases}
\]

(5.2)

where the exponent \( \beta \) is used to control the variance in the rainfall distribution at the particular location. The flexibility of this model lies in our ability to manipulate the autocorrelation, noise, and trend of the simulated rainfall series. In the analyses that follow, we first explore two limiting cases before fitting the model using data from Bamako in Mali and Mbita in Kenya. In the first limiting case, rainfall is described by a sinusoidal
trend with annual periodicity ($\mu_d = \mu_0 \sin(2\pi d/365)$) and no noise ($\sigma = 0$). In the second,
rainfall is subject to noisy and possibly autocorrelated variability ($\sigma > 0, \rho \geq 0$) yet with
no underlying trend ($\mu_d = \mu = \text{constant}$).

We fitted the model to the two case-study locations using 20 years of daily precipitation
data in each case (1st January 1995 to 31st December 2014, obtained from the European
Centre for Medium-Range Weather Forecasts, ECMWF). The parameter $\beta$ was fixed \textit{a priori} to give a good correspondence between simulated and actual rainfall series (Table
5.2, $\beta = 4$ for Bamako and $\beta = 5$ for Mbita). These values of $\beta$ were used to compute
$\{W_d\}$ by transforming the rainfall series (eqn. 5.2), where we assumed $W_d = 0$ on days
with no rainfall. To estimate the trend $\{\mu_d\}$, we next computed the mean value of $W_d$ for
each day of the year, averaged over the 20 year period. The resulting series $\{\mu_d\}$, was then
approximated by a Fourier series with $K$ harmonics,

$$\mu_d = \sum_{k=0}^{K-1} a_k \cos \left( \frac{2\pi dk}{365} \right).$$

(5.3)

$K = 4$ was chosen, which generated smoothed approximations of the $\{\mu_d\}$ series, which
produced rainfall series that had the most similar characteristics to the actual data (Table
5.2). Finally, $\rho$ and $\sigma$ were estimated by maximum likelihood.

5.3.4 HEG releases

The simulation environment is a 16 km$^2$ square area with periodic boundary conditions.
In all cases the landscape was run for a period of 50 days before the introduction of
mosquitoes, in order to allow the aquatic habitat dynamics to reach equilibrium. Similarly,
for all replicates the wild-type population was run for a period of at least 2 years before
the HEGs were introduced, in order to allow the population to reach equilibrium.

For the purely deterministic sinusoidal model, the releases occurred at four equally-
5.4. RESULTS

<table>
<thead>
<tr>
<th>characteristic</th>
<th>Bamako observed</th>
<th>Bamako simulated</th>
<th>Mbita observed</th>
<th>Mbita simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean wet day amount (mm)</td>
<td>4.90</td>
<td>4.65</td>
<td>9.13</td>
<td>9.76</td>
</tr>
<tr>
<td>Standard deviation of wet day amount (mm)</td>
<td>4.8</td>
<td>4.61</td>
<td>13.30</td>
<td>13.90</td>
</tr>
<tr>
<td>Median amount (mm)</td>
<td>0.00</td>
<td>0.04</td>
<td>3.35</td>
<td>1.82</td>
</tr>
<tr>
<td>P(dry tomorrow</td>
<td>dry today)</td>
<td>0.87</td>
<td>0.90</td>
<td>0.65</td>
</tr>
<tr>
<td>P(wet tomorrow</td>
<td>wet today)</td>
<td>0.67</td>
<td>0.66</td>
<td>0.85</td>
</tr>
<tr>
<td>P(wet)</td>
<td>0.28</td>
<td>0.23</td>
<td>0.70</td>
<td>0.60</td>
</tr>
<tr>
<td>Mean wet spell length (days)</td>
<td>7.65</td>
<td>9.58</td>
<td>2.87</td>
<td>3.10</td>
</tr>
<tr>
<td>Mean dry spell length (days)</td>
<td>2.99</td>
<td>2.91</td>
<td>6.61</td>
<td>4.68</td>
</tr>
<tr>
<td>Standard deviation of monthly totals (mm)</td>
<td>56.78</td>
<td>50.88</td>
<td>103.14</td>
<td>102.47</td>
</tr>
<tr>
<td>Maximum daily precipitation (mm)</td>
<td>110.75</td>
<td>42.26</td>
<td>158.95</td>
<td>222.83</td>
</tr>
<tr>
<td>Lag 1 autocorrelation</td>
<td>0.41</td>
<td>0.46</td>
<td>0.29</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 5.2: A comparison between rainfall data and model simulated data for Bamako, Mali, and Mbita, Kenya. For simulated series, the values represent averages across 100 replicates.

5.4 Results

5.4.1 Constant environment

In a constant environment, a released HEG will either spread to fixation or fail to establish (North, Burt, and Godfray, 2013). If fixation occurs, the load will either suppress or
extinguish the population, depending largely on the intrinsic growth rate of the target
population (Deredec, Godfray, and Burt, 2011). Consider, for example, a release where
rainfall is constant so that the aquatic habitat density is $142 \text{km}^{-2}$ at all times. Fixation
occurred in half the simulations of this scenario, while the HEG became extinct with no
lasting consequences in the remaining half. Population extinction occurred in 80% of the
cases where the HEG became fixed, suggesting that the population intrinsic growth rate is
relatively small for the parameters we assume.

5.4.2 HEG dynamics are more variable in a seasonal environment

In an environment with an average aquatic habitat density of $142 \text{km}^{-2}$, yet with seasonal
variation in the density, a wider range of dynamics is possible (Figure 5.1). Now, fixation
occurred in 34% of simulations and invariably led to population extinction (Figure 5.1a).
The HEG became extinct with no lasting consequence to the population in 49% of simu-
lations (Figure 5.1b). In the remaining simulations, the HEG induced extinction despite
first becoming lost from the population (17% of simulations, Figure 5.1c) or, occasionally,
the HEG and wildtype males both remained in the population throughout the simulated
period of 6 years post release (0.4%, Figure 5.1d). A common feature of these dynamics
is a tendency of the HEG frequency to fluctuate in an annual yet unpredictable manner
(Figure 5.1 a-d grey lines). If a HEG fixes in a wet season, the population is likely to
become extinct in the following dry season because the scarcity of aquatic habitats exac-
erbates the genetic load. If a HEG fails to fix in a given wet season, it may become lost
from the population in the following dry season. In most simulations, the population will
recover after the HEG is lost, although this may take several years. In some simulations,
however, the remaining population is so impaired from the invasion that it too becomes
extinct during a subsequent dry season.
5.4.3 The start of the wet season is the optimal time for release

The relative probabilities of these outcomes are influenced by both the time of year of the HEG release, and by the average density of aquatic habitats through the year (Figure 5.2). Extinction is generally most likely if the HEG is released at the start of the wet season (blue lines) and least likely if the release is at the start of the dry season (mauve lines). Release time is important in seasonal environments because of differences in the probability that the HEG becomes fixed and, in particular, that this occurs in the year of release. Due to the tendency of HEG frequency to fluctuate, it is most effective to release a HEG when it is likely to start increasing in frequency. We find this coincides with the start of the wet season, when the target population is growing. An additional benefit of releasing at this time is a relatively high initial HEG frequency due to the target population being small. If the initial frequency is high, fixation may occur during the first wet season leading to population extinction in the following dry season. We thus find that the release size is generally more important in seasonal than constant environments, and large releases may compensate for sub-optimal release timing (Figure 5.3).

5.4.4 In seasonal environments eradication is easier to achieve if the average aquatic habitat density is low or high rather than intermediate

Irrespective of release time, we found the HEG is more likely to induce eradication in environments with a low rather than intermediate average aquatic habitat density, in contrast to the trend when conditions are constant (Figure 5.2a). This partly reflects a higher initial HEG frequency in the case of low average aquatic habitat density, because the target population is smaller, and thus a greater chance of fixation in the year of release. In addition, the possibility of extinction occurring after the HEG is lost from the population is more likely if the average aquatic habitat density is low (Figure 5.2b). If aquatic habitats
are generally rare, the mosquito intrinsic population growth rate will be particularly small
during the dry season; in these conditions a population that has been suppressed by the
impact of the HEG may randomly become extinct.

In contrast to this trend, we found the probability of extinction generally increases
with aquatic habitat density in environments where the average density is already high. A
higher density of aquatic habitats reduces the risk of a HEG becoming lost from the target
population, and thus promotes HEG establishment.

Our results suggest that the roles of initial HEG frequency and underlying vulnerability
of the target population are more important when aquatic habitats are generally rare,
while HEG loss in the dry season is important when aquatic habitats are more abundant.
However, the balance of these factors depends on release timing, and we expect that detailed
models will be required to discern their relative importance in specific scenarios.

5.4.5 Random variation in rainfall generally increases the probability
that a HEG will induce population extinction

We next investigate the role of random weather variation, by allowing variance in the dis-
tribution of daily rainfall (controlled by $\sigma$) and autocorrelation in the rainfall time-series
(controlled by $\rho$; Figure 5.4). In cases of low variance and low autocorrelation, we observed
only minor effects of the variability on the outcome of a HEG release (Figure 5.4a, blue ver-
sus black). The probability that a HEG induces population extinction increases, however,
if either the variance in rainfall is increased (purple versus blue) or if the rainfall time-series
becomes more autocorrelated (magenta versus blue). We found the probability of extinc-
tion is highest if the rainfall distribution has both high variance and high autocorrelation
(red).

Random variability in rainfall increases the probability of extinction because it increases
variability in the dynamics of the target population. If a HEG becomes fixed in a variable climate, the population is likely to become extinct in a subsequent drought (Figure 5.4d). Moreover, if the variability is great, the impact of a HEG may lead to population extinction even after the HEG becomes itself lost from the population (Figure 5.4b). These effects of variability are to some extent lessened by a reduced probability of fixation (Figure 5.4c).

Rainfall variability makes HEG establishment more difficult since a release may occur during drought, when there are few available aquatic habitats. For the parameter ranges investigated, we found the greater potential impact of a HEG in a variable environment almost always outweighed the greater difficulty in establishment.

Note that our analysis assumes the releases occur on random dates with respect to weather conditions, whereas a release programme may have some capacity to delay a release if conditions are particularly adverse. A more systematic strategy to dealing with an unpredictable climate may be to release multiple time-delayed batches of HEG-bearing mosquitoes rather than a single large batch, and the probability of fixation increases if this bet-hedging approach is taken (results not shown).

5.4.6 Releases during the first half of the wet season result in the highest chance of population extinction, although in some cases the time taken for extinction to occur may be lower by releasing HEGs during the dry season

We investigated HEG releases in Bamako, Mali and Mbita, Kenya, using simulated rainfall series after fitting the rainfall model to 20 years of data in each case. For each location a typical model-generated rainfall series gave a good visual correspondence to the actual series, and resulted in similar estimated aquatic habitat dynamics (Figure 5.5). A more formal comparison is given in Table 5.2, showing that the model is able to replicate a range
Simulations of a mosquito population in each of these locations typically end in population extinction during the dry season (Bamako) or a relative dry spell (Mbita) in the absence of permanent aquatic habitats. The issue of how mosquito populations survive dry periods remains controversial, with aestivation and long-range migration possible explanations (Dao et al., 2014). Since neither of these possibilities occur in our model, it is unsurprising that the simulations are unable to replicate dry season survival; particularly for the case of Bamako which experiences more annual variability. In order to investigate HEG releases (in Bamako and Mbita) without modelling the uncertain dry season ecology, we focus on environments with permanent aquatic habitats (representing riverside or lakeside pools) as well as ephemeral aquatic habitats subject to rainfall (Figure 5.6). To construct Figure 5.6, we assumed that aquatic habitats were 65% from variable sources for Bamako, and 80% variable for Mbita. In both cases these were the maximum percentage of variable aquatic habitats that could sustain wild-type populations at least 99% of the time, and still allowed a reasonable running time.

For Bamako, there is a clear benefit of releasing during the first half of the wet season, for landscapes with low, or high aquatic habitat densities (red and purple lines in top-left panel of Figure 5.6). The relatively large effect for the lowest aquatic habitat density landscape (purple line), was due to a combination of increased occurrence of extinction due to a HEG partially-invading the population, together with increased fixation probability (data not shown). This contrasted with the high aquatic habitat case (red line), where the incremental probability of extinction during June-August was due to increased fixation probability, and a higher chance that a fixed HEG induces extinction. The different mechanisms driving the extinction dynamics account for the slight difference in optimal time of year of release for the low and high aquatic habitat landscapes. Interestingly, landscapes
with moderate aquatic habitat densities did not show a noticeable effect of release time on the probability of population extinction, although compared to the other cases there was more variation in the time taken for extinction to occur (data not shown). By releasing in March, before the wet season begun, the average time for extinction occurred was c.1.5 years, whereas releasing in the previous month resulted in the HEG taking roughly twice as long to drive population extinction. This sensitivity of time of extinction to release month is due to the importance of a HEG establishing in its first wet season. March releases tended to produce a HEG that established during the coming wet season, whereas for February this establishment was delayed until the next year’s wet season; in both cases extinction often occurring during the first dry season following the wet.

The effect of time of year of HEG release was much less marked for Mbita, due to the absence of a marked dry season for this location (Figure 5.6 top-right panel). In contrast to Bamako, the release time which resulted in the shortest time to population extinction was the months before the short rains; March-April for medium-high aquatic habitat densities, and January for the landscape with the lowest aquatic habitat density considered. Although, again this effect was weaker than for Bamako, where a long dry season causes a more pronounced effect.

In the event of planning a HEG release, an important consideration will be the speed of effect, as well as the effect itself. It is preferable for the HEG to act as quickly as possible, both for operational reasons (the sooner to reduce the prevalence of malaria) and practical reasons (a faster acting HEG gives less opportunity for HEG resistance to evolve). Across the landscape qualities investigated here, there was a lack of consensus over the release time which resulted in the fastest-acting HEG. However, in the case of Bamako across some of the parameter range, this optimum did not coincide with the release time that resulted in the greatest chance of population extinction. For example, in the case of the
lowest aquatic habitat density landscapes, releases at the start of the wet season (in June) resulted in the highest likelihood of population extinction, whereas releases in the middle of the dry season (December-January) resulted in the quickest eradication. This effect can be explained by the fact that releases during the dry season are usually unsuccessful, but if the HEG does manage to fix, an extinction may follow soon after.
Figure 5.1: Four examples of the outcomes following a HEG release, for adult population densities of wild-type and HEGs (black and grey lines), and frequencies (red lines). Blue shading indicates wet seasons. Demographic parameters are set at values as described in Table 5.1. The mean aquatic habitat density was set at $\kappa = 142 km^{-2}$ with seasonal variation about this level as described in the text.
Figure 5.2: The impact of release timing (determined by line colouration) in an environment with an annual rainfall cycle, and in a constant environment (black), as a function of average aquatic habitat density. a. The unconditional probability that the HEG causes population extinction, which is decomposed ($a = b + c \times d$) into, b. the probability that a HEG that does not spread to fixation induces a population extinction, and the probability that a fixed HEG induces extinction ($c \times d$). The error bars indicate 95% confidence intervals based on a normal approximation with c.200 simulation runs. Demographic parameters are set at values as described in Table 5.1.
Figure 5.3: The effect of release size (the number of HEG bearing males introduced) on the probability of population extinction. The parameters and colouration follows Figure 5.2, with average aquatic habitat density = 128 $km^{-2}$.
CHAPTER 5. USING HEGS TO CONTROL VECTOR POPULATIONS

Figure 5.4: The impact of HEG releases in environments with stochastic variability in rainfall (coloured lines), and in a constant environment (black), as a function of average aquatic habitat density. The simulated rainfall series were constructed with low (blue and purple) or high (magenta and red) levels of autocorrelation, and with low (blue and magenta) or high (purple and red) levels of variance. The error bars indicate 95% confidence intervals based on a normal approximation with c.200 simulation runs. Demographic parameters are set at values as described in Table 5.1.
Figure 5.5: Rainfall data from Bamako, Mali and Mbita, Kenya (blue dots a. and c.), and corresponding time-series simulated from the rainfall model with fitted parameters (b. and d.) The red lines plot the aquatic habitat densities computed from the rainfall time-series by the method outlined in the text, with $\chi = 0.1$ and $\kappa = 100$. 
CHAPTER 5. USING HEGS TO CONTROL VECTOR POPULATIONS

Figure 5.6: Simulated HEG releases in Bamako and Mbita (a. and b.) The probability of extinction depends on the average density of aquatic habitats in the environment (determined by line colouration) and the month of release. Graphs c. and d. show the average monthly rainfall over the 20 years of data we collected for each location. Demographic parameters are set at values as described in Table 5.1. The error bars indicate 95% confidence intervals based on a normal approximation with c.200 simulation runs.
5.5 Discussion

The use of HEGs to suppress vector populations holds clear potential as a tool for suppressing malaria. Mathematical modelling plays a key role in developing this technology by enabling investigation of how HEGs will perform in mosquito populations that are shaped by complex factors such as genetic diversity and environmental heterogeneity. In this paper we have focussed on rainfall because this is a key factor underpinning the population dynamics of malaria vectors, yet one that has previously received little attention. Our results suggest that HEG-based programmes can be successful in a wide range of climates, although the timing of HEG-release may be important in environments where rainfall is highly variable.

In general, HEGs should be released at the start of wet periods when the target species is growing at the fastest rate, since this maximises the probability that the genetic construct will successfully establish to become fixed in the population. In highly seasonal environments, where the wet season is more predictable, applying this rule may be marginally easier, although often considerable uncertainty remains over the start of the rains. In environments where wet weather is less predictable, it may be harder to plan releases to coincide with wet periods. For environments with less predictable weather, a bet-hedging strategy of few and often releases will be preferable to occasional large releases. Furthermore, releasing small batches of mosquitoes relatively often may be easier from an operational standpoint, since it may be harder to rear large populations of modified insects in individual insectories. In addition, releases could be focussed on locations where there is permanent breeding habitat and so the target population is more stable. Although variability in rainfall will increase the difficulty of establishing a HEG, it may also increase the impact of HEGs that do establish. Our results show an established HEG will be more likely to extinguish a target population if rainfall is variable, because such a population
must persist through dry periods when the load imposed by the HEG is more likely to be fatal.

The strong selection pressure for the mosquitoes to evolve a defence against the HEG’s action needs to be taken into account when designing a release. This also depends on the specific nature of the mutation that could arise; to avoid spontaneous mutations, it is important to minimise the number of mosquitoes alive between release and extinction, whereas for HEG-induced mutations more importance should be given to the total number of transgenic mosquitoes. Time taken for eradication to occur is also of practical importance, since shorter periods would result in a faster decline in vectorial capacity, more rapidly reducing the malaria burden within a region. Here we used the time taken between release and extinction both as a proxy for the probability that resistance arises, and as a measure of suppression of malaria. However, future analyses should extend the model to explicitly include mutations (see Deredec, Burt, and Godfray, 2008 for a spatially-averaged model of resistance for HEGs). It is also possible that even before a HEG is released, a minority of the wild population may have a natural immunity to the HEGs action. In the event of a HEG release, the (strong) selective advantage given to these mosquitoes would cause their population to grow rapidly. Whether population extinction would result in this circumstance, however, would depend on the degree of immunity conferred to an individual, the heritability of the defence, as well as its prevalence and spread in the population.

The rainfall model developed in this paper has proved a useful tool for investigating effects of temporal variation on mosquito populations, and how this will influence HEG-based control. The strength of the model lies in its flexibility, whereby we have been able to reproduce a wide range of realistic patterns of rainfall with only a small number of parameters. The model may provide a useful framework for other studies that investigate the effects of temporal variation on vector biology, or more generally the relationships
between malaria and climatic variability.

In the Sahel region of Africa, as exemplified here by Bamako, seasonality is marked (Xue et al., 2004). Our model thus predicts that HEG releases occurring at the start of the wet season will be particularly effective in this region, resulting in extinction of the target population. It is beyond the scope of the current model to quantify the difference in impact between wet and dry season releases at a particular location given that many complex facets of mosquito behaviour are ignored. In particular, by assuming that mosquito behaviour is the same all year round, the model predicts that a population in Bamako will become extinct each dry season unless there are permanent aquatic habitats (even before HEGs are introduced). Although it is well documented that Anopheles gambiae s.l. mosquitoes become hard to find during a Sahelian dry season (Touré et al., 1998; Taylor et al., 2001; Dao et al., 2014), they invariably become common soon after each wet season begins, suggesting that total collapses do not occur.

The suggested early wet season release for HEGs can be contrasted with the results of modelling other biological methods of disease control. Wolbachia are symbiotic bacteria that are found in many insect species, and spread through mosquito populations via a mechanism known as cytoplasmic incompatibility (Yen and Barr, 1971), which is currently being explored as a way of controlling dengue fever (Moreira et al., 2009; Bian et al., 2010), carried primarily by the Ae. aegypti mosquito. The introduced bacteria has been demonstrated to cause a reduction in vectorial capacity, either by lowering mosquito mean lifetime (McMeniman et al., 2009), or by inhibiting disease transmission (Moreira et al., 2009; Bian et al., 2010). In order to overcome the deleterious effects of carrying the bacteria and spread to the wider population, it is necessary that the released frequency of infected females exceeds a threshold (Hancock, Sinkins, and Godfray, 2011). Hence, it could be optimal to release at times when population densities are lowest, although Hancock,
Sinkins, and Godfray, 2011 suggest that it may be preferable to release a larger number of mosquitoes during the wet season, in order to avoid the high mortality in the dry season. Similarly, in SIT (sterile insect technique) the load imposed on the population would be greatest during times of low population densities, although practically release timings are typically chosen to coincide with particular crop-growing seasons (Dyck, Hendrichs, and Robinson, 2006).

The question of how *Anopheles* populations survive the dry season remains controversial, and is currently dominated by two competing hypotheses. The first contends that adult mosquitoes undergo a dry season diapause (aestivation) by hiding in (currently unknown) shelters while the second contends that large numbers of adults migrate, possibly over large distances, soon after the onset of each wet season to repopulate suitable habitats (Lehmann et al., 2010; Adamou et al., 2011; Dao et al., 2014). The former hypothesis is supported by mark-recapture observations (remarkably, Lehmann et al., 2010 discovered a female *A. gambiae* s.s. still alive over seven months after being marked in the previous wet season) and experimental evidence (Adamou et al., 2011). It is possible that different members of the *Anopheles* s.l. complex use different strategies; recent Mali-based studies have suggested that *A. coluzzii* aestivates while the closely related *A. gambiae* s.s. persists by long distance migration (Adamou et al., 2011; Dao et al., 2014).

We chose not to incorporate either survival mechanism into our model due to the lack of consensus over which, if either, mechanism predominates, yet it is worth speculating how these behaviours might alter our conclusions. A target population that aestivates will essentially skip the dry season, and the population at the onset of a wet season will be a subset of the previous wet season population (those that survive aestivation). A HEG that becomes established during a wet season will therefore remain present in the following year, and so will continue to suppress the target population. The overall extent
of suppression may be less than predicted by the current model because the role of the dry season is reduced, yet the general dynamics will be similar. If dry season survival is due to long distance migration, on the other hand, the population will be replaced each year and a HEG that establishes in one season will be unlikely to persist. In this case, it may be necessary to release HEGs annually to guarantee persistent population suppression. However, it may also be possible to disseminate HEGs across a landscape by their release in a population that becomes a source of migration, thus knowledge of Anopheles gambiae s.l. migration patterns will be invaluable.

The clearly important role of dry season behaviour highlights the value of research in mosquito ecology to assessing HEGs as a vector-control intervention. Before HEGs might be deployed, it will be important to develop detailed site-specific models to predict the outcome of initial releases as accurately as possible. This work will clearly benefit from further research on the dry season behaviour of Anopheles gambiae s.l. in Sahelian Africa, and also from further study into the effects of weather in other locations where releases may take place. Field studies that shed light on the link between rainfall and breeding habitat will likewise be extremely helpful. Whilst in our modelling here rainfall is taken always to increase mosquito population sizes, it has been suggested that flood waters may actually wash away larvae, reducing the availability of viable aquatic habitats (Reiner et al., 2015). Temperature is known also to affect survival in a nonlinear way; if too hot or too cold, there is increased mortality (Craig, Snow, and Sueur, 1999), and has lead to temperature being used a predictor variable in generating maps of malaria prevalence as well as its inclusion in the many studies which investigate the climatic determinants of vector population dynamics (Reiner et al., 2015). Local geology and species-specific preferences for habitat may also be important considerations for planning releases, and further research into these areas would be beneficial.
5.6 Conclusion

I have demonstrated that mosquito population control using a driving homing endonuclease gene (DEG) that causes severe sex ratio distortion can be a viable strategy in a highly seasonal environment, and that the precise timing of releases during the year can be important in maximising success. In particular, our simulation results indicate that releases at the start of the wet season are preferable for maximising the probability that the wild-type mosquito population is extinguished. In practice, however, operational aspects of rearing and releasing batches of mosquitoes will influence the release timing which can be achieved. In harder to reach locations, it may not be possible to locally release DEGs and, therefore, it will be important to develop models for larger regions of the environment - which account for environmental and human ecological aspects of geography - which include models of weather, to determine optimal release strategies. How informative the result of these simulations are depends on the assumptions made about local mosquito ecology, however, and so emphasises the continued importance of field entomology.

I finish by noting that though the use of DEGs and other gene-editing strategies is a novel technology with great promise, it is essential that a rigorous governance framework is set up to regulate its use, both to avoid any negative health or environmental impacts and to assure a licence to operate by civil society in the countries where it may be employed.
Chapter 6

Conclusion

From Macdonald’s mathematical model for the transmission of malaria, we know that mosquito borne diseases are most sensitive to the duration of the lifespan of mosquitoes. This is because mosquitoes that live longer (1) have more opportunity to bite an infected host and become infected themselves, (2) have a greater chance of surviving the time it takes for the pathogenic organism to develop inside the mosquito, such that the mosquito becomes infectious, and (3) have more time to bite an uninfected human host and transmit the disease. In this thesis, I presented the results from meta-analyses of previously-published data from the two predominant approaches for estimating wild mosquito longevity: mark-release-recapture (MRR) experiments (in Chapter 2), and studies that involve the dissection of female mosquitoes (in Chapter 3). The evidence from both approaches suggests that the majority of mosquitoes live for less than 10 days. Whilst these estimates likely represent lower bounds on the true lifespan of wild mosquitoes (see below), their correspondence indicates that mosquitoes may live shorter lives than is generally believed. If we take the estimates as they stand, this implies that only a small minority of mosquitoes live long enough to vector the major mosquito borne diseases. If this is the case, then disease
transmission should be particularly sensitive to interventions that kill adult mosquitoes, since small changes in adult mortality would drastically alter the number of mosquitoes surviving the EIP to become infectious. Indeed, this was the main conclusion drawn from the Ross-Macdonald model used to justify indoor residual spraying during the Global Malaria Eradication Programme (Smith et al., 2012).

There are reasons to believe that both MRR and dissection-based experiments likely underestimate the lifespan of mosquitoes, however. In MRR experiments, the handling of mosquitoes may damage them, and since the experimental data I analysed did not include information about where the recaptures occurred, the estimates represent the combined effects of mortality and dispersal of mosquitoes out of the study area, resulting in conservative estimates of lifespan. In dissection-based studies, it is assumed that the count of dilations in a particular ovariole faithfully represents the number of gonotrophic cycles that a female mosquito has undergone; an ovariole where this is the case is termed diagnostic. Previous work indicates that the proportion of diagnostic ovarioles decreases as a mosquito gets older, due to a range of physiological processes that destroy the ovariolar dilations required to age mosquitoes by this method. This would mean that older mosquitoes require dissections of a large number of ovarioles to locate those few with the greatest number of dilations, which faithfully represent the gonotrophic age of a mosquito. Whilst the various entomologists who carried out the dissection experiments were probably aware of this issue, it is possible, and has been shown to frequently be the case (Hugo et al., 2008), that the parous age of many mosquito specimens is less than the true gonotrophic age. Overall, the assumptions of the two methods suggest that the estimates I present are best viewed as lower bounds on mosquito lifespan.

Despite the different assumptions made by each method, it is encouraging that those species and genera where observations were available from both sources had comparable
estimates of lifespan. Where discrepancies exist, however, I tentatively suggest that some of the differences in lifespan may be attributable to the variation in dispersal across mosquito species or genera. Specifically, because the estimates from MRR studies effectively include the rate of dispersal whereas dissection-based estimates do not, it is possible MRR study estimates may be relatively downward-biased. Indeed, across the majority of species and genera, the estimates from the dissection studies were greater. At the genus level, the greatest difference was obtained for *Culex* mosquitoes followed by *Anopheles* then *Aedes* (where the estimates from the MRR studies were actually higher). It is thought that aedine mosquitoes, which live in close proximity to humans often in urban settings, disperse the least during their lifetimes, explaining why this genus was associated with the smallest difference in estimates from both approaches. Further work is required, however, to confirm the variation in dispersal of species and genera.

Across both classes of experiment, there was evidence that mosquito lifespan varies between species and genera. This is perhaps not a surprising result given that many mosquito species are found in relatively confined geographic areas, and that mosquitoes will inevitably adapt to their local environment. It should be noted, however, that within most species, the variability in lifespan equalled, and in many cases, exceeded, the variability between species. This was particularly pronounced for those species, for example, *Ae. aegypti* and *A. gambiae*, which are found in a variety of geographies. On this basis, I argue that it is likely adaptation to local climate and other environmental pressures that drives the observed differences in lifespan. Whilst I investigated air temperature as a potential source of this variability (and found no systematic pattern), I believe that other factors, such as rainfall, may be more predictive of lifespan.

Laboratory colonies of mosquitoes are essential for studying mosquito borne diseases, yet we know that these mosquitoes and their environmental conditions are considerably
different to those in the wild. This is typified by the differences in mosquito lifespan for
laboratory and wild mosquitoes, where those in captivity regularly live for many multiples
of the estimated lifespan of their wild counterparts. Some of these differences are likely due
to biases in the methods used to estimate wild mosquito lifespan, yet these do not explain
the bulk of the discrepancy. Predation, the risk of mortality from blood-feeding and a lack
of food, amongst other factors present in wild environments, are likely major contributors to
these differences, yet our current knowledge of mosquito ecology is insufficient to conclude
which of these predominates. These differences in lifespan, I believe, highlight the issues
inherent with extrapolating conclusions drawn from laboratory colonies to the wild, and I
argue that a number of such published results (including those concerning age-dependent
mortality) may not generalise well to the field. Whilst more expensive, and constrained in
their use to those geographies that naturally support mosquito populations, I argue that
semi-field experiments - where many of the conditions are similar to wild environments
- will be of increasing importance for the study of mosquito borne disease. Just how
representative these semi-field environments are of the wild is currently unclear, however,
and further work is needed to justify their added expense. Despite this, it is encouraging to
see that these studies appear to be becoming more prevalent in recent years (see Hugo et al.,
2014 for an example of an experiment used to estimate mosquito mortality in Vietnam).
Additionally, it is important to have methods for determining characteristics of mosquito
ecology from wild caught specimens, and I believe that MRRs and dissection studies will
continue to play an important role in this area. Yet, it is still desirable to develop new
methodologies which are simpler to implement than these traditional approaches, and
currently there are a number of promising approaches which are subject to current research.

In Chapter 4, I discussed one such approach to estimating mosquito lifespan known as
near-infrared spectroscopy (NIRS). In this approach, light is shined through the head and
thorax of mosquitoes, and the adsorption across a range of wavelengths is measured. As a mosquito gets older, its biochemistry changes, causing changes to its infrared spectrum. Previous work has demonstrated that these changes to the spectrum are regular enough that machine learning algorithms can be trained on them and used to predict the ages of individual mosquitoes. In this chapter, I applied a more recently developed machine learning algorithm than has been used previously, to this task. The initial results from using this approach showed a significant improvement in the average errors at the individual level, but (like previous work) produced bias estimates of lifespan, with over-prediction for younger mosquitoes and under-prediction for older individuals. Since this bias was systematic, by modifying the machine learning algorithm, I was able to obtain an unbiased estimator of individual mosquito lifespan.

A potential use of NIRS in the field could be to detect changes in the average lifespan of the local mosquito population in response to vector control measures, such as the deployment of insecticide treated bednets and the indoor spraying of insecticide. In Chapter 4, I demonstrated that NIRS applied to a modest number of caught mosquitoes can reliably detect differences in the average lifespan of a population, along those age ranges thought to be realistic. Further, because the bulk of the errors in population lifespan were due to sampling error (the random variation in mosquitoes caught by collections), NIRS was found to be relatively insensitive to inflations in the individual prediction errors, so long as these estimates remain unbiased.

There is reason to believe, however, that the errors obtained from applying NIRS in practice will not be of qualitatively different form to those I have assumed here. In Chapter 4, I demonstrated that NIRS calibrated on data from one experiment cannot predict the ages of individual mosquito specimens scanned in another. There is insufficient data to determine the source of the inter-experiment heterogeneity, but it is possible that this is
due to differences in the spectrometer, the method used to preserve the caught mosquitoes, and the genetics of the specimens, amongst other factors. If the bulk of this variation is due to methodological differences, then it should be possible to establish protocols which minimise this effect, and NIRS could be straightforwardly used to survey wild populations. If the source of the variation is due to non-experimental heterogeneities, such as species and geography, then each application of NIRS will require its own calibration dataset composed of local mosquitoes (possibly from wild larvae raised in a laboratory) of known ages. Whilst this necessity would diminish the worth of NIRS, it could, nevertheless, be a potentially useful tool for field entomology.

In Chapters 2 and 3, I compared the fit of a number of different models for mosquito mortality to data. One of these models - the exponential distribution - assumes a constant (i.e. age independent) risk of mortality, whereas the other five models I used allowed for a mortality risk that increased with age. For the MRR experiments, the exponential model fitted the data as well as the other five models across each of the species in the dataset, indicating that we cannot reject the hypothesis that wild mosquitoes do not experience an elevated mortality risk as they age. For the dissection experiments, the results were more mixed. In nearly half of the species, the exponential model fit the data as well as or better than the best performing of the five models which incorporate senescence. In the rest of the species, the exponential did not provide as good a fit to the data as the best of the other five, however, although in some of these cases, the exponential model still performed better than one or more of the models with age-dependence. Collectively, I believe these results indicate that mosquito senescence, should it occur in the wild, is likely a subtle effect; considerably smaller than suggested from experiments on laboratory colonies (for example, see Styer et al., 2007). Presumably, in the wild, environmental factors (as discussed above) result in the premature death of mosquitoes before they get to ages where they experience
senility. Again, experiments which indicate the relative importance of these factors would allow a better understanding of these differences and, perhaps, suggest novel methods for vector control. Due to the subtlety of mosquito senescence, I argue that differences in the modelled impact of vector control using the Ross-Macdonald model and those incorporating senescence may be less major than is sometimes suggested.

Widespread use of existing vector control methods, including most notably, insecticide treated bednets and the spraying of insecticides indoors, has likely been a major contributor to the recent gains against malaria (Bhatt et al., 2015). The growing prevalence of insecticide resistance amongst the major disease vectors, however, threatens to halt and, perhaps, reverse these trends (Ranson et al., 2011). Similarly, the spread of parasites resistant to artemisinin - an important drug used to treat human malaria - in Asia (Ashley et al., 2014; Imwong et al., 2017), is extremely concerning and suggests a greater emphasis on finding new ways to control the disease. A recently proposed approach involves using genetic drive technologies to spread traits which impose a fitness load on a mosquito vector population, leading to declines in its numbers and, potentially, its elimination. In Chapter 5, I investigate the use of one such technology, known as a homing endonuclease (HEG), which has been engineered to bias the sex ratio of mosquito progeny towards males. As a so-called Y-drive HEG spreads through the population, it should become increasingly composed of males, leading to an eventual population collapse when the few remaining females die. Previous modelling work has investigated the release of these HEGs in both panmictic (Burt, 2003; Derede, Burt, and Godfray, 2008; Derede, Godfray, and Burt, 2011) and spatially heterogeneous (North, Burt, and Godfray, 2013) environments. In this chapter, I build on this work by incorporating a rainfall model which is linked to the availability of aquatic habitats in a previously-published spatial model of mosquito ecology (North, Burt, and Godfray, 2013), and using it to investigate the release of HEGs in environments
with temporal heterogeneity. This work demonstrates that increased heterogeneity can
detrimentally impact the ability of a HEG to establish in a population, but that this effect
is more often than not compensated for by the population’s elevated vulnerability. This
means that, across much of parameter space, temporal variation in the environment can
actually boost the performance of a HEG release, meaning it more often leads to population
elimination. This suggests that HEGs, or more generally driving endonuclease mechanisms
(DEGs), show great potential for vector control in the field. Encouragingly, a number of
DEGs have been successfully inserted into laboratory populations of mosquitoes (Wind-
bichler et al., 2011; Gantz et al., 2015; Hammond et al., 2016) and a consortium called
Target Malaria lead by Austin Burt (the first to propose using HEGs for vector control)
has been actively involved in evaluating the potential use of such technologies for vector
control.

The model I used in Chapter 5 assumes that resistance to DEGs does not arise over
the time scales it takes for a DEG to suppress mosquito population densities. We know,
however, that there exist mechanisms that could give rise to resistant strains of mosquitoes.
For example, it is possible that the DNA sequence that a DEG targets may not be con-
served across all individuals in a wild population. After the release of a DEG, any in-
dividuals lacking this recognition sequence would, hence, have a selective advantage, and
likely spread through the population. Similarly, chromosomal repair can occur by non-
homologous mechanisms such as end joining, which could generate alleles which did not
contain the requisite recognition sequence. Regardless of the source of resistant alleles, its
spread will be dictated by the other fitness costs associated with the mutation. Modelling
work has determined the characteristics of the engineered DEG which determine the pop-
ulation dynamics associated with resistance in non-spatial environments (Deredec, Burt,
and Godfray, 2008; Godfray, North, and Burt, 2017), and attempts to model the spread of
a DEG with resistance in spatially explicit environments are now being actively pursued. It is hoped that the modelling can inform the work of biological teams who are responsible for engineering the DEGs and inserting them into the mosquito genome, leading to a technology that avoids a number of sources of resistance.

Others have argued that the gaps in our knowledge of mosquito ecology are significant barriers to disease eradication (Godfray, 2013; Ferguson et al., 2010a), and it is my hope that this thesis helps to slightly narrow the gap associated with mosquito survival. It is clear, however, that there are important areas of mosquito ecology and behaviour about which we know little, and I now speculate on those areas which I believe may prove fruitful in the future.

Whilst it is generally accepted that there is substantial variation in the ecology of different species of mosquito, I believe that there is less recognition of the variation within a given species. For example, the use of insecticide treated bednets and the indoor spraying of insecticides for malaria control is based on the stereotypical perspective that vectors bite indoors (at night) and rest on the inside walls of houses (Ferguson et al., 2010a). It is known, however, that endophilic species may sometimes feed outside, and may do so in response to interventions which target domestic behaviours (Pates and Curtis, 2005; Reddy et al., 2011). Whether this behaviour represents phenotypic plasticity or has underlying genetic causes, is currently not well known. Our view of putting mosquitoes in ecological and behavioural ‘boxes’, whilst historically useful, is now, perhaps, detrimental to the cause. In order to build mathematical models which can be used to accurately predict disease transmission following interventions, I believe that we need to move away from regarding a species’ behaviour as a discrete point in ecological space towards a distribution, whose spread is informed by experimental work. Characterising the covariances in mosquito ecological space between vector species is also of crucial importance. For
example, when interventions suppress endophilic species, this gives a selective advantage to exophilic vectors, which in some cases, may be equally capable of vectoring disease (a good example being the major malaria vector *A. arabiensis*). Understanding how vector species interact with predators and prey throughout their various life stages is also, of importance (Godfray, 2013). An intervention which effects one part of an ecosystem could cause an ecological disturbance resulting in another vector species arising in prominence (Ferguson et al., 2010a). Unless we have a reasonable understanding of these intra- and inter-species interactions, I believe, that computational models of mosquito ecology will likely overestimate the effects of interventions.

An element of mosquito ecology that I argue is currently poorly understood is the dispersive behaviour of mosquitoes. As well as being an important element of models of disease transmission, spatial dispersal is of particular importance to the design of DEG programmes (Godfray, North, and Burt, 2017), where it is hoped that modified mosquitoes in one area will seed the invasion of DEGs into other locations. Classic metapopulation theory suggests that the population dynamics which result from a DEG release are likely to be highly sensitive to the assumptions made about the spatial dispersal kernels of mosquitoes (Hanski, 1998). The spatial kernels determine the connectivity of patches (idealised discrete areas of the landscape) and, hence, determine the rate at which wild-type populations are invaded by DEGs, and also the rate at which empty patches are colonised by either DEGs or wild-type individuals. The meta-population dynamics also depends on any differences in dispersive fitness which exist between wild-types and DEG individuals. Clearly, if wild-type individuals are considerably fitter than the DEGs then it will be more difficult to eliminate wild-type populations. The spatial dispersal kernel not only affects the rate of DEG spread but also how resistant alleles, should they arise, will spread through wild populations. Again, the outcome of the invasion of the resistant allele will depend on its
relative fitness versus wild-type and DEG individuals.

Mark-release-recapture experiments remain our best source of information on the dispersal of mosquitoes (see, for example, Guerra et al., 2014). Yet, it remains unclear how to design these experiments to capture the requisite information, and the numbers of individuals that would need to be released to best estimate the tails of the kernels; likely to be of particular importance to the design of interventions. I am currently working with Ace North on a spatial model of MRR experiments which, we believe, may help to answer some of these questions. Of course, it may prove to be the case that the numbers of individuals required to probe the tails of the spatial kernels is beyond the remit of traditional experimental approaches. Godfray, 2013 has suggested that an alternative marking methodology would be to spray a large area of a landscape with harmless chemical elements, such as Rubidium, which have been demonstrated to be taken up by larvae and are retained into the adult stage (Wilkins et al., 2007).

Another part of mosquito ecology that is crucial to the spread of a DEG is the dry season behaviour of anopheline mosquitoes in those climates (for example, the Sahel region of Africa) where there are distinct wet and dry seasons. There are a number of hypothesis concerning how these mosquitoes seem to appear soon after the first rains fall, including aestivation (Lehmann et al., 2010), long-range migration (Dao et al., 2014) and ovipositing in the few water bodies that persist through the dry season. Yet, to date, there is not strong consensus on which of these, if any, can explain the observed population dynamics.

As well as the existing gaps in our knowledge of mosquito ecology, our manipulation of our environment brings new challenges. We are currently in the midst of the greatest period of urbanisation in human history. In 1910, 10% of the world’s population resided in cities, whereas today there are more urbanites than non-urbanites, and the percentage of the population living in cities is expected to continue growing in the near future (UN,
2002). This unprecedented, and widespread change to our way of life dramatically changes the ecology of an environment (Grimm et al., 2008), and likely exerts potent evolutionary forces on species that dwell within cities (Partecke, Schwabl, and Gwinner, 2006). Others have argued that urbanisation of cities changes the ecological landscape enough to affect the transmission of mosquito borne disease (although differ in their proposed directionality of the effect on disease transmission; Keiser et al., 2004; Hay et al., 2005), but it is possible that the selective pressure exerted on mosquitoes may have deleterious consequences for us. In particular, it is possible that hitherto zoophilic mosquitoes might adapt their biting behaviour towards human sources - a ubiquitous source of food in cities. Alternatively, anthropophilic species that are accustomed to living in natural environments might adapt to oviposit and breed in close proximity to humans (for example, typical of the behaviour of *Ae. aegypti*). If an African anopheline species that liked urban environments and favoured biting humans arose, then projections that suggest malaria transmission will decrease (see, for example, (Hay et al., 2005)) would be wildly incorrect, and a new, urban, era of malaria might begin.

Malaria and many other vector borne diseases have plagued humans from a time predating civilisation. Since the work of Manson, Ross, Lazear, and Reed, amongst others at the turn of the twentieth century put mosquitoes at the centre of transmission for these horrific illnesses, we have had considerable successes and, unfortunately, notable failures against these flying emissaries. The recent declines in malaria cases in those parts of the world most affected have, finally, given us some cause for optimism. It is important, however, that we do not now rest our laurels on the clay appendages of the opposition. There are many reasons to believe that malaria and other mosquito borne diseases will continue to affect those vulnerable many throughout the course of this century. A sinister recent development is the discovery in Cambodia of a single dominant artemisinin-resistant form
of *P. falciparum*, which has now developed resistance to piperaquine - the drug the country uses as a first line treatment for malaria (Imwong et al., 2017). The tragedy of the recent Zika epidemic for those worst affected parts of the world is unarguable and is, perhaps, the most serious infectious disease to threaten the Western world since HIV/AIDS invaded the United States in the 1980s. Even though this disease doesn’t kill, we are just beginning to understand its impact and it hints at the real risk posed by emergent *Aedes*-borne diseases. If any good is to come from all this human suffering, it is through our - the West’s - re-acquaintance with the ills carried by our mosquito foe which, I hope, leads to greater funds being allocated to its study.
Chapter 7

Appendix

7.1 A mathematical model of senescence

I now present a simple model to provide insight into the effect of mosquito senescence - that is age-related death - on vector control efforts. The model is similar in nature to that presented in Bellan, 2010, although here I use a continuous age structured model and explicitly include a larval population (incorporating density dependence). I use this model only to undertake an analysis of the statics of the system, in other words, when the populations are at their steady states, as this provides ample insight into the impact of an age-dependent mortality on vector control efforts. I leave the analysis of the dynamic behaviour of the system for future work.

Vectorial capacity is a measure of the transmission potential of a mosquito population and is often used to assess the impact of vector control efforts (Smith and McKenzie, 2004). Vectorial capacity is similar in nature to $R_0$ and represents the expected additional human infections that will be generated each day from a single infected human. This quantity is derived from the product of four terms: (1) the emergence rate of mosquitoes $\epsilon$, (2) the
average number of human bites for a single mosquito throughout its lifetime, $S$, representing the initial biting of an infected human (for vectorial capacity calculations we assume the mosquito always becomes infected), (3) the probability that a mosquito survives through the entomological inoculation period (EIP), $Pr(A > n)$, so that it can transmit sporozoites to an uninfected human in a subsequent bite, that (4) is again represented by $S$,

$$C = \epsilon \times S \times Pr(A > n) \times S = \epsilon S^2 Pr(A > n). \quad (7.1)$$

If mosquitoes do not experience the effects of old age then individual lifespan is exponentially distributed and, hence, $Pr(A > n) = e^{-ng}$, where $g$ is the (age-independent) force of mortality, and $n$ is the duration of the EIP. Further, for an exponential lifespan distribution, it is straightforward to calculate $S$, the average number of human bites a mosquito takes in its lifetime; it is given by the product of the average lifespan of a mosquito, equal to $\frac{1}{g}$, and the human biting rate, $a$, which measures the average number of bites per mosquito per day,

$$S = \frac{a}{g}. \quad (7.2)$$

If we assume that the mosquito population is at its steady state,

$$\frac{dM}{dt} = \epsilon - g\bar{M} = 0, \quad (7.3)$$

meaning $\epsilon = \bar{M}g$, where $\bar{M}$ is the steady state population density. Using this result we obtain an alternative expression for vectorial capacity, which is the same as that derived from the Ross-Macdonald model of malaria transmission (Macdonald, 1957),

$$C = \frac{\bar{M}a^2e^{-ng}}{g}. \quad (7.4)$$
To derive an equivalent quantity for a population that experiences an age-dependent mor-
tality risk, I consider a continuous time age-structured model of a population of adult
mosquitoes (see, for example, Murray, 2002, for an introduction to these types of model).
To enable an assessment of the effects of vector control approaches employing larvicide, I
also include a larval population in the model. To keep things simple, I assume that the
larval population does not experience age-dependent mortality and, hence, can be well-
described by an age-averaged model. I assume that the larval population compete for
resources and represent this density-dependence by a logistic term. The model system is
of the form,

\[
\frac{dL(t)}{dt} = fM(t)\left(1 - \frac{L(t)}{\kappa}\right) - \eta L(t) \tag{7.5}
\]

\[
\frac{\partial m(A,t)}{\partial A} + \frac{\partial m(A,t)}{\partial t} = -\mu(A)m(A,t) \tag{7.6}
\]

where \(L(t)\) is the density of the larval population, \(m(A,t)\) is the density of the adult
population of age \(A\) at time \(t\), \(f\) is the fecundity of an adult mosquito (here I consider only
a population of females and implicitly assume the presence of a male population), which is
assumed to be independent of age, \(M(t) = \int_0^\infty m(A,t)\,dA\) is the overall density of the adult
population, \(\kappa\) is the carrying capacity of the larval population, \(\eta\) is the rate of recruitment
of the larval population into the adult stages, and \(\mu(A)\) is the force of adult mortality, that
can depend on age. To close this system, I assume that those larvae recruited into the
adult population enter it at age \(A = 0\) (here \(A\) measures adult age),

\[
m(0,t) = \eta L(t). \tag{7.7}
\]
Considering the populations at their steady states yields the following relations,

\[
\frac{\dot{M}}{M} \left(1 - \frac{L}{\kappa}\right) = \eta L \tag{7.8}
\]

\[
\frac{d\tilde{m}(A)}{dA} = -\mu(A)\tilde{m}(A), \tag{7.9}
\]

where \( \tilde{L}, \tilde{m}, \) and \( \tilde{M} \) represent steady state quantities. The second of these expressions can be solved by separation of variables to yield the steady state age structure of the adult population,

\[
\tilde{m}(A) = \tilde{m}(0)e^{-\int_0^A \mu(\tau)d\tau} \tag{7.10}
\]

\[
= \eta L e^{-\int_0^A \mu(\tau)d\tau} \frac{S(A)}{S(A)} \tag{7.11}
\]

where \( S(A) \) is the survival function, which indicates the proportion of the adult population surviving till age \( A \). The steady state adult population density is calculated by integrating expression (7.11),

\[
\tilde{M} = \eta L \int_0^\infty S(A)dA \tag{7.12}
\]

\[
= \eta L \bar{D} \tag{7.13}
\]

where the second line follows because the integral of the survival function across all ages is the average lifespan of the population (Cox and Oakes, 1984), \( \bar{D} \). By substituting expression (7.13) into expression (7.8), we can solve for the steady state larval population density,

\[
\tilde{L} = \kappa \left[1 - \frac{1}{Df}\right], \tag{7.14}
\]
7.1. A MATHEMATICAL MODEL OF SENESCENCE

where we obtain a non-zero steady state larval density if \( \bar{D}f > 1 \), in other words, the average number of offspring produced by a mosquito during its lifespan exceeds one - sufficient to replace itself. The steady state number of larvae recruited into the adult population per day is, hence, given by,

\[
\tilde{m}(0) = \eta \kappa \left[ 1 - \frac{1}{\bar{D}f} \right].
\]  

(7.15)

I now follow Styer et al., 2007 and calculate the vectorial capacity of each age cohort (a collection of individuals of the same age). Its derivation follows the same approach that we used to calculate the overall vectorial capacity, and results in the following expression,

\[
C(A) = a^2 \tilde{m}(A) e(A) Pr(t > A + n|t > A)
\]

(7.16)

where \( e(A) \) is the expected remaining lifespan of a mosquito that has reached age \( A \), and \( Pr(t > A + n|t > A) \) is the probability that this mosquito survives for a subsequent duration of time equal to the EIP (\( n \)).

In what follows, I suppose that adult mosquito lifespans can be described by the Gompertz distribution (see Chapter 2), which has an age-specific force of mortality equal to,

\[
\mu(A) = de^{A\lambda}
\]

(7.17)

where \( d \) and \( \lambda \) are constant parameters that affect the shape of this distribution. When \( \lambda = 0 \), the force of mortality is a constant (equal to \( d \)), independent of age, and we obtain a exponentially-distributed lifespan distribution (blue line in Figure 7.1A). For an exponential distribution for lifespans, the distribution of mosquito ages is also exponentially distributed (with the same rate parameter), and the survival curve is also an exponential
function (blue line in Figure 7.1B). As $\lambda$ is increased, the mosquitoes are more sensitive to 
the effect of old age, and the lifespan distribution becomes increasingly truncated towards 
older mosquitoes (orange and green lines in Figure 7.1A), and few mosquitoes survive much 
longer than the mean lifespan (orange and green lines in Figure 7.1B).

\[ \begin{align*}
A. \text{lifespan distribution} & \\
B. \text{survival curve} & \\
\end{align*} \] 

Figure 7.1: The probability distribution for mosquito lifespan (left) and the 
survival curves for these mosquito populations (right) as a function of the 
Gompertz distribution’s $\lambda$ parameter. In all cases, the populations have a mean 
lifespan of 8 days.

I now use expression (7.16) to illustrate how the vector capacity is affected by age-
dependent mortality. To do this, I compare a population that experiences no age-related 
mortality (left column of Figure 7.2) with one that experiences significant age-related decay 
(right column). In particular, for the senescent population, I specify $\lambda = 0.2$, which 
corresponds to the lifespan distribution and survival curves shown by the orange lines in 
the left and right panels of Figure 7.1, respectively. I assume the mean lifespan (8 days) 
and EIP (10 days) are the same in both cases. Expression (7.16) shows that the age-specific
7.1. A MATHEMATICAL MODEL OF SENESCENCE

vector capacity is the product of three things (in what follows, I assume $a = 1$ since this
does not affect the reasoning): (1) the density of mosquitoes at a given age (top row of
Figure 7.2), (2) the residual life expectancy for a mosquito of that age (second row from
top) and, (3) the probability that the mosquito subsequently survives for a duration equal
to the EIP (second row from bottom). The bottom row in Figure 7.2 shows the age-specific
vectorial capacity (solid lines) and the overall vectorial capacity (shaded areas) for both
populations.

The most significant differences between the elements of vectorial capacity for the two
populations are the residual life expectancy (second row from top of Figure 7.2) and the
probability of surviving the EIP (second row from bottom). For the population that does
not experience senescence, these two quantities are independent of age (left column), due
to the memoryless property of the exponential distribution. Whereas for the senescent
population (right column), both of these quantities decrease with age, with the probability
of surviving the EIP being particularly sensitive. The result of this age-dependence is that
the vectorial capacity is considerably lower for the senescent population: for the parame-
ters I use here, the population experiencing age-independent mortality has a transmission
intensity over three times that of the senescent case.

I now derive expressions for the overall vector capacity for a population experiencing
age-related mortality. To do this, we note that $Pr(t > A + n|t > A)$ is a conditional
probability and, hence, can be derived from the law of conditional probability,

$$Pr(t > A + n|t > A) = \frac{Pr(t > A + n)}{Pr(t > A)}$$

(7.18)

$$= \frac{S(A + n)}{S(A)}.$$  

(7.19)

Using expression (7.16), this allows us to derive another expression for the age-specific
vectorial capacity,

\[ C(A) = a^2 \tilde{m}(A) e(A) \frac{S(A+n)}{S(A)} \]
\[ = a^2 \tilde{m}(0) S(A+n) e(A). \]

(7.20)

(7.21)

This result can be interpreted intuitively: \( S(A+n) e(A) \) is the average number of days a mosquito of age \( A \) will last after it has survived the EIP, meaning that the above expression represents the average number of infections generated per day from mosquitoes of age \( A \). The overall vectorial capacity is then just the integral of this expression,

\[ C = a^2 \tilde{m}(0) \int_0^\infty S(A+n) e(A) dA. \]

(7.22)

If the population experiences no effect of old age, then the above reduces to,

\[ C = \tilde{m}(0) \frac{a^2 e^{-gn}}{g} \int_0^\infty e^{-gA} dA \]
\[ = \frac{a^2 \tilde{m}(0) e^{-gn}}{g^2} \]
\[ = \frac{\tilde{M} a^2 e^{-gn}}{g}, \]

(7.23)

(7.24)

(7.25)

and we recapitulate the Ross-Macdonald result given in expression (7.4). (To obtain the final line I used \( \tilde{m}(0) = \tilde{M} g \).)

I now use the model to compare how interventions that reduce adult lifespan (for example, insecticide treated nets or pyrethroid spraying) or larval survivorship (for example, the destruction of larval habitat or chemical larvicide) affect the vectorial capacity. I assume that both classes of interventions act in age-independent fashion (see Hancock, Thomas, and Godfray, 2009 for an analysis of the use of fungal pesticides that target older
mosquitoes), with an elevated risk of mortality given by $\zeta$, for adults, and $\chi$, for the larval stage. For the adult population, this amounts to using a Gompertz-Makeham model (see Chapter 2) for the lifespan distribution, which has an age-specific force of mortality equal to,

$$\mu(A) = de^{A\lambda} + \zeta. \quad (7.26)$$

To consider the effect of larval interventions, I change the larval population dynamic equation to,

$$\frac{dL(t)}{dt} = fM(t)\left(1 - \frac{L(t)}{\kappa}\right) - \eta L(t) - \chi L(t), \quad (7.27)$$

where $\eta$ represents the rate at which larvae emerge into adults and $\chi$ is the rate at which they die due to the treatment.

I consider the effect of these interventions on the steady state values of the vectorial capacity. In Figure 7.3, I illustrate how the vectorial capacities of a population that does not experience senescence (left column) and a senescent population of the same mean lifespan ($\lambda = 0.2$, right column) are affected by an intervention that increases the adult mosquito mortality rate by $\zeta = 0.1 \text{ day}^{-1}$. In both cases, the elevated adult death rate causes a decline in the mosquito population density (top row of Figure 7.3). For the population where age does not affect mortality risk, the increased adult mortality significantly reduces the residual life expectancy of mosquitoes of all ages (left panel, second from top of Figure 7.3), whereas for the senescent case, the reduction in life expectancy is greatest for the youngest mosquitoes although not as large in magnitude (right panel, second from top).

The effects on the life expectancy are mirrored by the probability that a mosquito of a given age survives the EIP (second row from bottom of Figure 7.3) where, again, the
impact is greater on the non-senescent population. Overall, this means that the impact on
the disease transmission is greatest for the population where mortality is age-independent
(bottom row of Figure 7.3): for the parameters I use here, the vectorial capacity is reduced
by about 80% for the population that does not experience age-related death, and 68% for
the senescent population.

I next compare the sensitivity of the vectorial capacity for the same two populations
as the intervention effectiveness is varied. Figure 7.4A shows the percentage reduction
in vectorial capacity for the population without senescence (blue line) and the one with
age-dependent mortality (orange line). In all cases, the proportional reduction in disease
transmission is lower for the senescent population. The greatest difference between the two
populations in the magnitude of the effect occurs when $\zeta \approx 0.06$, when the intervention
reduces the vectorial capacity for the population without senescence by about 16% more
than for the other case. However, I also believe it is worth noting that, in both cases,
the reduction in disease transmission that can be achieved by killing adult mosquitoes is
considerable for even modestly-effective interventions.

To calculate effects of larvicide on the vectorial capacity, I need only consider its effect
on the larval population density (which affects vectorial capacity through expression (7.7)),
because it does not impact adult survivorship. For the system described by expressions
(7.6) and (7.27), the steady state larval density is given by,

$$\tilde{L} = \kappa \left[ 1 - \frac{\eta + \chi}{f_\eta D} \right], \quad (7.28)$$

where the population can sustain itself so long as $\eta(f\bar{D} - 1) > \chi$, in other words, the average
density of newly-emergent adults exceeds the rate at which larvae are killed. Expression
(7.28) results in a steady state density of newly-emergent adults equal to,

\[
\tilde{m}(0) = \eta \kappa \left[ 1 - \frac{\eta + \chi}{f \eta D} \right]. 
\]

Using this expression, we can determine the percentage reduction in vectorial capacity by the following expression (obtained from the ratio of the steady state density of newly-emergent adults after the intervention to that before it),

\[
r = \frac{\chi}{\eta (f D - 1)}. 
\]

I illustrate how this function varies as a function of the strength of the intervention \( \chi \) in Figure 7.4B. Since this relationship is independent of the specific form of mortality for the adult population, I show only a single line in this plot. For the parameter values that I use here, this intervention has relatively minor effects on disease transmission compared to those that impact adult mortality (Figure 7.4A). This echoes the result of Macdonald, 1957, whose equation for \( R_0 \) partly motivated the programme initiated in the mid-twentieth century to kill adult mosquitoes by DDT spraying (Packard, 2007).
Figure 7.2: The vector capacity at a given age (bottom row) for a population without senescence (left column) and another with age-dependence ($\lambda = 0.2$, right column) is calculated by taking the product of the density of mosquitoes of that age (top row) multiplied by the remaining life expectancy of that mosquito (second row from top) and the probability it survives the EIP (second row from bottom). In both cases, I assume that the life expectancy of a mosquito is 8 days, the EIP is 10 days, $\kappa = 1$, $f = 5$, and $\eta = 1$. 

Figure 7.3: The impact of an intervention that kills adult mosquitoes ($\zeta = 0.1$, orange lines) versus the pre-intervention quantities (blue lines) on the components of vectorial capacity (top three rows) and the vectorial capacity itself (bottom row) for a population without senescence (left column) and another with age-dependent mortality (right column). In both cases, I assume that the life expectancy of a mosquito is 8 days, the EIP is 10 days, $\kappa = 1$, $f = 5$, and $\eta = 1$. 
Figure 7.4: The sensitivity of vectorial capacity to elevated adult mortality (A) for populations without senescence (blue line) and with senescence (orange line) and the same sensitivity for elevated larval mortality (B, black line). In all cases, I assume that the life expectancy of a mosquito is 8 days, the EIP is 10 days, $\kappa = 1$, $f = 5$, and $\eta = 0.1$. 
7.2 Studies included in the MRR meta-analysis

The following is the subset of studies from the original Guerra et al., 2014 database which were used in the MRR meta-analysis of Chapter 2: Marini et al., 2010; Baber et al., 2010; Lacroix et al., 2009; Freitas, Eiras, and Oliveira, 2008; Midega et al., 2007; Maciel-De-Freitas, Codeco, and Lourenco-De-Oliveira, 2007; Elizondo-Quiroga et al., 2006; Ba et al., 2005; Fabian et al., 2005; La Corte Dos Santos, Forattini, and Burattini, 2004; Watson, Saul, and Kay, 2000; Harrington et al., 2001; Tsuda et al., 2001; Muir and Kay, 1998; Touré et al., 1998; Quinones et al., 1997; Costantini et al., 1996; Trpis, Häusermann, and Craig Jr, 1995; Jensen and Washino, 1994; Fernandez-Salas, Rodriguez, and Roberts, 1994; Jaal and MacDonald, 1992; Rodriguez et al., 1992; Chiang et al., 1991; Jensen and Washino, 1991; Eldridge and Reeves, 1990; Macdonald, Sebastian, and Tun, 1968; Pumpuni and Walker, 1989; Charlwood and Bryan, 1987; Charlwood and Alecrim, 1989; Birley and Charlwood, 1989; Arredondo-Jiménez, Rodríguez, and Washino, 1998; Hii, Birley, and Sang, 1990; Renshaw, Service, and Birley, 1994; Milby and Reisen, 1989; Freitas, Codeco, and Oliveira, 2007; Loong et al., 1990; Nelson and Milby, 1980; Freitas et al., 2006; McDonald, 1977; Curtis and Rawlings, 1980; Rawlings and Davidson, 1982; Conway, Trpis, and McClelland, 1974; Reisen, Mahmood, and Parveen, 1979; Nelson et al., 1978; Rawlings et al., 1981; Sempala, 1981; Takagi et al., 1995; Buei et al., 1980; Eyles and Bishop, 1943; Ordóñez González et al., 2001; Pant and Yasuno, 1973; Charlwood, Graves, and Marshall, 1988; Reisen et al., 1982; Nayar, Provost, and Hansen, 1980; Carnevale et al., 1979; Eyles, Sabrosky, and Russell, 1946; Reisen et al., 1984; Charlwood, Graves, and Birley, 1986; Trpis and Hausermann, 1975; Lutwama and Mukwaya, 1994; Wada et al., 1969; Takken et al., 1998; Abdel-Malek and Abdel-Aal, 1966; Valerio et al., 2012; Zetek, 1915; Takagi et al., 1995; Yasuno, Rajagopalan, and Laüreque, 1975; Eyles and Cox, 1943; Germain, Hervé, and Geoffroy, 1974.
7.3 Studies included in the dissection study meta-analysis

The following is a list of studies included in the dissection study meta-analysis of Chapter 3: Catangui, 1971; Charlwood and Wilkes, 1979; Chang et al., 1991; Charlwood, Vij, and Billingsley, 2000; Edalat et al., 2015; Barros, Honorio, and Arruda, 2011; Lines, Wilkes, and Lyimo, 1991; Magesa et al., 1991; Lines et al., 1991; Forattini et al., 1996; Samarawickrema, 1967; Samarawickrema, 1968; Barros et al., 2007; Charlwood et al., 1985; Chandra, Seal, and Hati, 1996; Hoc and Wilkes, 1995; Vythilingam et al., 1997; Russell, 1986a; Chadee and Ritchie, 2010; Ebsary and Crans, 1977; Charlwood, 1980; Shriram and Krishnamoorthy, 2011; Mahmood and Reisen, 1981; Samarawickrema, Sone, and Cummings, 1987; Uttah et al., 2013; Ramaiah and Das, 1992; Buei and Ito, 1982; Chadee et al., 1995; Chandra, 2008; Foll and Pant, 1966; Kanda, Joo, and Choi, 1975; Mala et al., 2014; Shalaby and World Health Organisation, 1962; World Health Organisation, 1960; Reisen, Mahmood, and Parveen, 1980; Detinova, 1962; Smith and Kurtz, 1994; Jayanetti, Wijesundera, and Amaresinghe, 1987; Ch et al., 2013; Pant and World Health Organisation, 1962; Ghosh, Mandal, and Chandra, 2010; Mahanta et al., 1999; Schlein and Müller, 2012; Penilla et al., 2002; Schlein and Müller, 2015; Snow and Boreham, 1978; Detinova, 1968; Nathan, 1981; Charlwood, Dagoro, and Paru, 1985; Schlein and Müller, 2008; Beier et al., 2012; Müller and Schlein, 2005; Surendran et al., 2006; Müller et al., 2010; Mendis et al., 1998; Wilkes, Matola, and Charlwood, 1996; Chatterjee and Chandra, 2000; Müller et al., 2010; Tuchinda et al., 1969; Chan, 1971; Aigbodion et al., 2011; Qualls et al., 2015; Reisen, Mahmood, and Azra, 1981; Gillies and Wilkes, 1972; Gillies and Wilkes, 1965; Kay, 1979; Hitchcock Jr, 1968.
7.4. Studies included in the gonotrophic cycle duration meta-analysis

The following is a list of studies included in the gonotrophic cycle study meta-analysis of Chapter 3: Gillies and Wilkes, 1965; Sheppard et al., 1969; Meillon, Sebastian, and Khan, 1967; Pant and Yasuno, 1973; Germain, Hervé, and Geoffroy, 1974; Lowe et al., 1973; Fernandez-Salas, Rodriguez, and Roberts, 1994; Buei et al., 1980; Rawlings and Curtis, 1982; Mori and Wada, 1977; Sempala, 1981; Reisen et al., 1983; Birley and Charlwood, 1989; Suzuki, 1978; Charlwood and Bryan, 1987; Charlwood, Graves, and Birley, 1986; Charlwood and Wilkes, 1979; Charlwood et al., 1995; Chang et al., 1991; Edalat et al., 2015; Barros, Honorio, and Arruda, 2011; Samarawickrema, 1968; Samarawickrema, 1967; Charlwood et al., 1985; Chandra, Seal, and Hati, 1996; Russell, 1986a; Chadee and Ritchie, 2010; Mahmood and Reisen, 1981; Ahumada, Lapointe, and Samuel, 2004; Kenawy, 1991; Rajagopalan, 1980; Chandra, 2008; Scholl, Porter, and Defoliart, 1979; Mala et al., 2014; Afrane et al., 2005; Gillies, 1953; Quinones et al., 1997; Rúa et al., 2005; Delatte et al., 2009; Arredondo-Jiménez, Rodríguez, and Washino, 1998; Wong et al., 2014; Ijumba, Mosha, and Lindsay, 2002.
Bibliography


[6] Y. A. Afrane et al. “Effects of microclimatic changes caused by land use and land cover on duration of gonotrophic cycles of *Anopheles gambiae* (Diptera: Culicidae) in western Kenya highlands”. In: *Journal of Medical Entomology* 42.6 (2005), pp. 974–980.


[118] M. Dubrulle et al. “Chikungunya virus and Aedes mosquitoes: saliva is infectious as soon as two days after oral infection”. In: PLoS one 4.6 (2009), e5895.


[125] A. Elizondo-Quiroga et al. “Gonotrophic cycle and survivorship of Culex quinquefasciatus (Diptera: Culicidae) using sticky ovitraps in Monterrey, northeastern Mex-


[219] T Kanda, C. Joo, and D. Choi. “Epidemiological studies on Malayan filariasis in an inland area in Kyungpook, Korea, 2: The periodicity of the microfilariae and the
bionomics of the vector [Brugia malayi, Anopheles sinensis].” In: Mosquito News (1975).


L. Pauling et al. “Sickle cell anemia, a molecular disease”. In: (1949).


[370] W. Samarawickrema, F. Sone, and R. Cummings. “Seasonal abundance, diel biting activity and parity of *Aedes polynesiensis marks* and *A. samoanus* (Grünberg)


320

BIBLIOGRAPHY


