High-resolution geospatial analysis for disaster risk management

Steven Louis Rubinyi

Christ Church College



Thesis presented for the degree of Doctor of Philosophy

Environmental Change Institute

School of Geography and the Environment

University of Oxford

Trinity Term, 2020

Abstract

Rapid advancements in Earth observation, localised data collection, and modelling techniques are revolutionising global understanding of disaster risk. If properly validated and appropriately used, these emerging techniques for high-resolution geospatial analysis can help developing countries overcome challenges pertaining to the first priority of the *Sendai Framework for Disaster Risk Reduction 2015-2030 (SFDRR)*: understanding disaster risk. This thesis focuses on three potential breakthroughs in risk information for developing countries: (i) mapping population exposure; (ii) demonstrating more granular insights into population vulnerability; and, (iii) measuring the extent of flood events.

For mapping population exposure, a simplified equation is presented that separates masking variables and density variables. The equation focuses on two distinct objectives: (i) increasing the precision and accuracy of a strong binary mask to limit the spatial extent of settled populations; and, (ii) improving modelling techniques of ancillary variables to account for variations in density within an identified spatial extent. A critical assessment of population modelling techniques that heavily rely on derived characteristics of the built environment as a masking variable is also presented. The assessment confirms that density variables directly relating to the binary masking variable are effective; however, their application requires careful consideration due to its propensity to amplify both positive results and errors. The research provides evidence in support of hybrid techniques towards global population modelling.

For demonstrating more granular insights into population vulnerability, a novel methodology is presented for generating spatially referenced synthetic households by combining population synthesis techniques for small area estimation with dasymetric population modelling. The methodology results in an output vector layer of households represented as points with unique structural and socioeconomic attributes which can be applied towards a better understanding of population characteristics, particularly those which are geographically influenced and may not conform to administrative boundaries linked to aggregated census statistics. A case study of Dumuria Upazila, Bangladesh, validates the methodology and provides an example of how georeferenced surveys can be applied towards error assessments in data-scarce regions, which remains a challenge in global population modelling.

Lastly, for measuring the extent of flood events, a novel methodology applies freely available synthetic aperture radar (SAR) data for flood detection over large geographic areas immediately following a cyclone hazard event. The methodology results in enhanced precision, demonstrated in coastal districts of Bangladesh heavily affected by 2019 Cyclone Fani. The case study also demonstrates how high-resolution geospatial analysis can supplement locally reported estimates of damages by providing quicker estimates and a means for identifying potential bias from self-reporting. Furthermore, insights into the socioeconomic characteristics of potentially affected households provide a layer of information not covered by post-disaster needs assessments.

Key Words: geospatial analysis; disaster risk; built environment; population modelling; population synthesis; flood detection

Acknowledgments

This research has been made possible with support from the American Friends of Christ Church, the REACH Programme (funded by UK Aid from the UK Department for International Development), the Global Facility for Disaster Reduction and Recovery (GFDRR), and the World Bank.

I would like to acknowledge the quality of mentorship, guidance, and feedback received over the years from my supervisor, Professor Jim Hall. For your support, I am ever grateful. To Professor Rob Hope, Professor Douglas Gollin, Dr Katrina Charles, Dr Sonia Hoque, and Professor Andrew Tatem, each of whom assessed my research through the transfer of status, confirmation of status examination, or viva, thank you.

Thank you as well to the countless colleagues and friends who have supported me throughout this journey. Lastly, thank you to my family: to my parents for instilling the virtue of relentless curiosity; to my siblings for providing inspiration; and to my partner for always providing motivation when it was most requisite.

Publications

The work presented in Chapters 3 through 5 has been submitted or accepted to the following peer-reviewed journals:

1. Rubinyi, S., Blankespoor, B., and Hall, J. W. The utility of built environment geospatial data for high-resolution dasymetric global population modeling. Submitted to *Computers, Environment and Urban Systems* on April 12, 2020. (Chapter 3).
2. Rubinyi, S., Hall, J.W., and Gussenbauer, J. Spatially referenced synthetic households for high-resolution population modeling. Submitted to *Journal of Geographical Systems* on December 25, 2019. (Chapter 4).
3. Rubinyi, S., Goldblatt, R., Gussenbauer, J., Hall, J.W., Kowarik, A., Mannix., J., Bottoms, B. High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone. Submitted to *Weather and Climate Extremes* on April 24, 2020. (Chapter 5).

Table of Contents

Abstract i

Acknowledgments iii

Publications iv

Table of Contents v

List of figures viii

List of tables xi

1. Introduction 1

1.1. Background and rationale 1

1.2. Aims and objectives 8

1.3. Structure of thesis 10

2. The challenge of understanding disaster risk 12

2.1. Introduction 12

2.2. Definitions and key concepts 13

2.3. Disaster risk reduction as a global priority 18

2.4. Towards a high-resolution understanding of global disaster risk ... 23

2.5. Key breakthroughs in disaster risk information for developing countries 33

2.6. Conclusion 45

3. The utility of built environment geospatial data for high-resolution dasymetric global population modelling 47

3.1. Introduction 48

3.2. Methodology 55

3.3. Cast Study: New York City 59

3.4. Discussion 71

3.5. Conclusions 77

3.6. Data Statement 79

4. Spatially referenced synthetic households for high-resolution population modelling 81

4.1. Introduction 82

4.2. Methodology 90

4.3. Case study: Dumuria Upazila, Bangladesh 97

4.4. Discussion 110

4.5. Conclusions 118

5. High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone 120

5.1. Introduction 121

5.2. Methodology 129

5.3. Results 142

5.4. Discussion 154

5.5. Conclusions 158

6. Concluding Remarks 160

6.1. Thesis summary 160

6.2. Primary research contributions 163

6.3. Practical Applications 165

6.4. Future Research 167

Appendix A: Co-author Paper Contribution Statements 171

Bibliography 178

List of figures

Fig. 1 The Risk Triangle (adapted from Crichton (1999)) 14

Fig. 2 Simplified example of risk to populations (left) and risk to assets (right) 16

Fig. 3 Areal interpolation techniques for a source area with a population ys of 80..... 59

Fig. 4 Census counties (boroughs), tracts and blocks at the intersect between Manhattan, Queens and Brooklyn boroughs..... 64

Fig. 5 Land use (labelled by land use code) and building footprints (labelled by building height in meters) layers with 30-meter resolution gridded overlay near Columbus Circle, Manhattan 65

Fig. 6 New York City population density utilising a residential building footprint binary mask and volume density variable downscaled from census tracts... 66

Fig. 7 Mapped results of the seven methodologies (20x downscaling) at the intersect between Manhattan, Queens and Brooklyn boroughs..... 67

Fig. 8 %RMSE and %MAE for each methodology at 20x and 8,000x downscaling 72

Fig. 9 %RMSE and %MAE for each borough with all building footprint binary mask and residential building footprint binary mask 73

Fig. 10 Coefficient of variation and %MAE of census blocks within census tracts at 20x downscaling using GAM smoother..... 77

Fig. 11 Map of Study Area Unions..... 99

Fig. 12 Workflow for Simulation of the Population with SimPop..... 102

Fig. 13 Combined distribution of household size within study area 103

Fig. 14 Combined synthetic population pyramid for study area 103

Fig. 15 30-meter resolution grid cells overlain by building footprints (left), 30-meter resolution modelled gridded households (centre), and modelled household points (right) in Sahas Union, Khulna District 105

Fig. 16 Point features for spatially referenced synthetic households and sample from accompanying attribute table for selected features (in light blue) 106

Fig. 17 Age distribution of surveyed vs synthetic populations 107

Fig. 18 Point features for spatially referenced synthetic households overlaid by point features for surveyed households.....108

Fig. 19 Selected *mauzas* within Bhandar Para Union are designated as "NF BBS 2011" per the 2011 National Census and are listed as unpopulated 109

Fig. 20 Linear model of households observed vs households predicted with 95 percent confidence interval and results of accuracy tests for 27 *mauzas* within Sahas and Sarappur Unions.....110

Fig. 21 Overlay of areas that have been inundated at least once on the full Dartmouth Flood Observatory record (1985-2010) and spatially referenced synthetic households.....115

Fig. 22 Study area comprised of (west to east) Satkhira, Khulna, Bagerhat, Pirojpur, Barguna, and Patuakhali districts.....131

Fig. 23 Difference Image DIFF (left) and Smoothed Difference Image SDIFF (right) 134

Fig. 24 Point feature class of predicted household locations spanning Patharghata, Barguna Sadar, and Amtali *upazilas* in Barguna District 137

Fig. 25 Point feature class of synthetic households and associated attribute table of selected features in Barguna District 142

Fig. 26 Synthetic households (dark points) and Cyclone Fani flood extent from May 4, 2019 (in blue) spanning sections of Bagerhat and Pirojpur Districts 143

Fig. 27 Comparison of observed (blue) vs. UNOSAT (grey) flooded area ... 146

Fig. 28 Comparison of observed (blue) vs. UNOSAT (hatch pattern) flooded area 147

Fig. 29 District-wise comparison of observed vs. UNOSAT flooded area (left) and percentage of total district area (right) 148

Fig. 30 District-wise comparison of modelled potentially flooded households vs. NDRCC reported damaged households (left) and percentage of total district population (right) 150

Fig. 31 Total modelled potentially flooded households by variable (left) and proportional distribution of modelled by district (right) 152

Fig. 32 Difference from the expected distribution for variables of modelled potentially flooded households 153

List of tables

Table 1. Deaths, wind speed, and severity index of major cyclones (>1,000 number of deaths) in Bangladesh (1960-2020).....5

Table 2. Key activities listed in the SFDRR for understanding disaster risk at national and local levels.....22

Table 3. Prominent open access gridded population data sets that cover multiple countries by first year of release.....35

Table 4. *Prominent gridded population data sets that cover multiple countries by first year of release*.....51

Table 5. New York City population, land area, and population density by borough.....61

Table 6. Data Layers for New York City Analysis.....62

Table 7. *New York City land use codes and descriptions*.....65

Table 8. Borough-level disaggregation accuracy assessment (8,000x downscaling).....69

Table 9. Tract-level disaggregation accuracy assessment (20x downscaling) 70

Table. 10 Classification of census blocks by population density (persons per square kilometre).....75

Table 11. Classification of census blocks by population density.....75

Table 12. Area, Population, and Household Data for Bhandar Para, Sahas, and Sarappur Unions.....98

Table 13. Data Layers Used in Dumuria Upazila Analysis.....101

Table 14. Upazilas with at least 10 percent observed flooded area.....144

Table 15. Unions with at least 20 percent observed flooded area.....145

1. Introduction

1.1. Background and rationale

Each year there are over 20,000 earthquakes recorded globally, yet only a small fraction have both the necessary severity and proximity to vulnerable populations and assets to result in severe losses (Hayes et al., 2017; Munich Re, 2018). Those that do can be devastating. On January 12, 2010, a magnitude 7.0 earthquake struck Haiti, resulting in the loss of 230,000 lives and an estimated $8 billion of damages (Kolbe et al., 2010; Cavallo et al., 2010). Less than two months later, a magnitude 8.8 earthquake struck Chile (approximately 500 times stronger than the Haiti earthquake) yet resulted in fewer than 1,000 deaths and $30 billion of damages (Kovacs, 2010; Miranda et al., 2012).

Why the divergence? Despite its much larger relative strength, the epicentre of the Chile earthquake occurred further from major population centres and at a greater depth than the Haiti earthquake (EERI, 2010). The nearest large city to the epicentre in Chile, Concepción (population of 200,000), was located approximately 70 miles away. In Haiti, the epicentre of the earthquake occurred a mere 10 miles from its largest city and capital, Port-Au-Prince (population of 3 million) (Vanholder et al., 2011). Haiti's exposed populations and assets were also much more vulnerable; Chile established a strict building code following a devastating 1985 earthquake while Haiti had no national building code and no means of checking building safety (Kovacs, 2010). As a much poorer country, Haiti had fewer resources available to invest in disaster mitigation and preparedness *ex ante* (before the event) and in response and

recovery *ex post* (after the event). Correspondingly, Chile's Gross Domestic Product (GDP) in 2010 amounted to approximately 33 times that of Haiti, meaning that the higher gross damages endured by Chile represent a much smaller proportional share of its GDP (18 percent) when compared to Haiti (110 percent) (Kovacs, 2010; World Bank, 2010a).

While natural hazards are inevitable, they are also, to an extent, predictable because they are inherently linked to the spatial characteristics of the natural environment. Chile is positioned directly along the "Ring of Fire", a zone surrounding the Pacific Ocean where approximately 90 percent of the world's earthquakes occur, including the five strongest earthquakes of the past 100 years (Kious et al., 1996; Scholz, 2019). Haiti sits directly along the Enriquillo–Plantain Garden fault zone and has experienced several severe earthquakes in its past, albeit at a lower frequency than Chile (prior to the 2010 earthquake, the last major earthquake in Haiti was in 1842). This holds true for other categories of natural hazards as well: volcanoes cluster along subduction zones and hotspots, landslides and avalanches occur in high-slope ranges, sinkholes in areas with carbonate rocks, floods along rivers and coasts, and cyclones in the tropics. Although natural hazards differ significantly in terms of their recurrence intervals (e.g. seasonal flooding versus caldera-forming volcanic eruption), the size of the affected area (e.g. landslide versus cyclonic storm surge), and their lead time (e.g. earthquake versus drought), the probability of a hazard affecting a specific geographic area can generally be estimated (Dilley et al., 2005; Dolce et al., 2012).

The 2010 Haiti and Chile earthquakes also highlight how the relative impact of hazard events can differ dramatically depending on the exposure and vulnerability of populations and assets within the built environment. The likelihood that populations and assets are affected by a hazard event (e.g. loss of life, injury,

destruction, damage) is defined as disaster risk and is characterised by the interplay between hazard, exposure, and vulnerability (UN General Assembly, 2016). Disaster risk management applies policies and strategies which aim to prevent new disaster risk, reduce existing risk, and manage residual risk. Resilience, which refers to the ability of those exposed to a hazard to resist, absorb, accommodate, adapt to, transform, and recover from its effects, has also become a key focus for disaster risk management in recent years (UN General Assembly, 2016). Enhancing resilience of individual, assets, and services serves to reduce the long-term impact of events beyond initial damages to consider wide-ranging consequences such as impact on livelihoods and well-being (Hallegatte et al., 2019).

Examples of investments in disaster risk management include flood control infrastructure, early warning systems, contingency planning, emergency operations centres, building code enforcement, and capacity building of key personnel. When joined together, investments in disaster risk management pay dividends. Strengthening infrastructure systems in developing countries presents a $4.2 trillion investment opportunity over the lifetime of new assets, where for each $1 spent countries can realise $4 in benefit due to fewer disruptions and reduced economic impacts (Hallegatte et al., 2019).

An example of the effective application of disaster risk management policies and investments is in the Bangladesh Coastal Zone (BCZ). The BCZ is an area demarcated by the Coastal Zone Policy (Ministry of Water Resources, 2005), that covers 19 districts, 47,201 square kilometres, and 710 kilometres of coastline and is home to approximately 42 million people (Bangladesh Bureau of Statistics, 2012). The BCZ is highly vulnerable to tropical cyclones and subsequent storm surges, which are projected in climate models to increase in frequency and intensity in

Bangladesh due to climate change (Huq et al., 2010; Uddin et al., 2019). 62 percent of coastal lands are less than three metres above sea level, while 83 percent are less than five metres above sea level (Bangladesh Water Development Board, 2012).

A series of risk reduction measures were put in place in the decades following the 1970 Bhola Cyclone, which resulted in the loss of 500,000 lives. The 1970 Bhola Cyclone is considered the deadliest tropical cyclone, and one of the worst natural disasters on record (Lopez-Caressi et al., 2013). Further investments in disaster risk reduction were intensified across the BCZ following the 1991 Bangladesh Cyclone, which took the lives of approximately 138,000 people (Shaw et al., 2013). This included investments in early warning systems, coastal embankments, and cyclone shelters, but also a coordinated effort by the Government of Bangladesh to better manage disasters through the establishment of the Disaster Management Bureau (Haque et al., 2012). The impact of these and subsequent investments is clear: the 2007 Cyclone Sidr resulted in less than one percent of the deaths caused by the 1970 Bhola Cyclone and three percent of the deaths caused by the 1991 Bangladesh Cyclone, despite being comparable in strength (Table 1).

Table 1. *Deaths, wind speed, and severity index of major cyclones (>1,000 number of deaths) in Bangladesh (1960-2020)*

Year	Deaths	Wind Speed (km/h)	Severity Index	Landfall Location
1960	8,119	210	5	Noakhali, Chittagong
1961	11,466	146	5	Meghna Estuary
1963	11,520	203	5	Noakhali, Chittagong
1965	20,152	210	5	Barisal, Noakhali
1970	500,300	223	6	Bhola, Meghna Estuary
1985	11,069	154	5	Noakhali
1988	9,590	162	5	Cox's Bazar (Teknaf Upazila)
1991	138,958	225	6	Noakhali, Chittagong
2007	4,234	250	6	Barguna

Source: Haque et al. (2012) and Dasgupta et al. (2014)

The United Nations has adopted the *Sendai Framework for Disaster Risk Reduction 2015-2030* (SFDRR) to help facilitate "the substantial reduction of disaster risk and losses in lives, livelihoods and health and in the economic, physical, social, cultural and environmental assets of persons, businesses, communities and countries" (UN General Assembly, 2015). The SFDRR highlights four priorities for action, the first of which is "understanding disaster risk". The idea behind this priority is simple: extensive high-quality data form the foundation for deriving accurate information, which enables more informed decision-making across the four phases of the disaster risk management cycle: (i) mitigation; (ii) preparedness; (iii) response; and, (iv) recovery (Khan et al., 2008; Othman et al., 2014). Data are critical for informing

granular risk assessments; investment planning of lifeline, critical, and risk-mitigating infrastructure; coordination of disaster response; and, rapid identification of recovery needs.

Unfortunately, high costs related to collecting and maintaining extensive high-quality data have hindered the ability of developing countries to reap the benefits which a firm understanding of disaster risk can help to unlock (Deichmann et al., 2011). Early warning systems for hydro-meteorological hazards have traditionally required large investments in monitored and remotely sensed data as inputs for intensive modelling before forecasting can lead to dissemination, decision-making, and eventually value (Lazo et al., 2008). High costs for local data collection paired with low capacity for data management, maintenance, and analysis lead to limited knowledge of exposed populations and assets. These challenges can be amplified in rapidly urbanising developing countries, where informal settlements are often constructed in hazardous areas such as steep hillsides in Sao Paolo and floodplains in Dar Es Salaam (Baker, 2012). Granular socioeconomic data, which can inform vulnerability, is particularly difficult to acquire due to coverage and privacy concerns, as well as high costs associated with national censuses (Arzberger et al., 2004; Gomez, 2017). The 2020 United States National Census is estimated to cost $15.6 billion, a sum larger than the total government expenditures of many developing countries (United States GAO, 2019; International Monetary Fund, 2020). Despite these challenges, a proliferation of new data sources is enabling disaster risk to be assessed quicker, cheaper, and easier at ever-higher spatial resolutions, providing the basis for more effective disaster risk management.

Earth observation, extended data collection methods, and high-resolution modelling are providing more detailed and granular insights into our understanding of

hazard (Twele et al., 2016; Mosavi et al., 2018; Mangalathu et al., 2020), exposure (Stevens et al., 2015; Lloyd et al., 2017; Pesaresi, 2018; Ehrlich et al., 2018), and vulnerability (Sedda et al., 2015; Jean et al., 2016; Utazi et al., 2020). MODIS, Landsat, and Sentinel satellite constellations all offer freely available data with global coverage, which has led to a broad transformation in availability of baseline satellite data for analysis regarding both the natural and built environment (Yuan et al., 2020). The trend is set to continue with the regular launch of new satellites carrying advanced sensors and enabling more frequent scene captures. For example, two additional Sentinel-1 satellites featuring C-band synthetic aperture radar (SAR) data will be launched in 2021, and the NASA-ISRO (NASAR) satellite, once launched in 2022, will provide open SAR data 4-6 times per month at 5-10-metre resolution. Concurrently, the rapid advancements in unmanned aerial vehicles (UAVs) paired with sensor technologies have led the cost of more localised remote sensing (e.g. lidar) to plummet (Ackerman, 2016). Crowdsourced geospatial data, critical for ground-truthing and training of data sets for use in machine learning models targeting image classification, object detection, and semantic segmentation, are also becoming easier to collect with the proliferation of GPS-enabled smartphones and through grassroots support of platforms such as OpenStreetMap (Diou et al., 2018). Additionally, emerging data sources which focus on human mobility and activity (e.g. call data records, smartphone location data, credit card transaction data, etc.) are increasingly being made available by private companies at resolutions which aim to protect anonymity and can be used to provide more dynamic insights into population characteristics relevant for disaster risk management (Steele et al., 2017; Birkin, 2019; Dujardin et al., 2020).

Improving upon techniques for high-resolution modelling, which leverage the wealth of new data and advancements in modelling techniques, are critical for advancing our knowledge of disaster risks (Le Cozannet et al., 2020). This is particularly true for developing countries, where large information gaps can hinder efforts to manage disasters effectively (Li et al., 2019). Thus, there is a corresponding and continuous need to evaluate the accuracy, precision, and applicability of increasingly high-resolution data towards managing disaster risks. This thesis addresses a research gap targeting three emerging areas of high-resolution insights with significant potential to advance our understanding of disaster risk in developing countries: (i) global population modelling for high-resolution insights into exposure; (ii) population synthesis of household structural and socioeconomic characteristics for a more granular understanding of vulnerability; and, (iii) synthetic aperture radar from satellite remote sensing for rapid and precise estimates of flood extent.

1.2. Aims and objectives

This thesis aims to evaluate the effectiveness of emerging techniques and propose new techniques for high-resolution geospatial analysis applied to disaster risk management. This is important for advancing our understanding of novel forms of high-resolution spatial insights and to translate those insights into actions that can help developing countries overcome challenges pertaining to the first priority for action outlined in the SFDRR: understanding disaster risk. The overall aim of this thesis is broken down into three research questions, each of which targets a separate component of disaster risk: exposure, vulnerability, and hazard. Each research

question is addressed by achieving its corresponding objectives, and is the focus of an individual chapter in this thesis:

1. How accurate are emerging techniques for high-resolution global population modelling that are reliant on precise mapping of the built environment?
    - Assess the accuracy of a range of binary masking variables reliant on precise mapping of the built environment (land areas, building footprints, residential building footprints)
    - Assess the accuracy of a range of density variables reliant on precise mapping of the built environment (building footprint areas, building volumes)
2. How can more granular and spatially nuanced insights be provided to inform population vulnerability in data scarce environments?
    - Develop a methodology for generating spatially referenced synthetic households by combining population synthesis techniques for small area estimation with population modelling
    - Generate a population of spatially referenced synthetic households in a data scarce environment and validate the results
3. How can satellite remote sensing techniques be applied to better understand the impacts of natural hazards and supplement disaster risk management practices in developing countries?
    - Delineate flooded areas from a tropical cyclone hazard event using satellite remote sensing techniques
    - Compare the structural and socioeconomic characteristics of modelled flooded and non-flooded households to locally reported estimates

- Determine where modelled high-resolution insights can provide added value across the disaster risk management cycle

1.3. Structure of thesis

This thesis follows the guidelines put forth by the School of Geography and the Environment, University of Oxford, for submission of an article-based thesis. Chapter 2 presents a review of the role that technological and methodological advancements have played in advancing our understanding of exposure, vulnerability, and hazard, and the applicability of those advancements towards informing disaster risk management. Chapter 2 also presents the theoretical and methodological background from which the research questions and corresponding objectives for this thesis were derived, covering key topics such as: (i) the continual improvement of models and outputs for global population modelling; (ii) population synthesis methods for small area estimation; and, (iii) flood event detection using satellite remote sensing.

Chapter 3 investigates recent advancements in global population modelling, a key input for understanding population exposure in data scarce environments. Several recent efforts utilising advanced methods to disaggregate census population counts from administrative units have produced gridded population data at up 30- and 10-metre resolution by relying on precise spatial inputs of the built environment. Chapter 3 tests the accuracy of a range of methods that progressively increase the precision of a binary masking variable (land areas, all building footprints, residential building footprints) and of density variables relating to the binary masking variable (building footprint areas, building volumes). The assessment is applied in New York City, where large spatial heterogeneities exist across confined geographic areas.

Chapter 4 addresses the challenge posed by limited granularity and spatially nuanced insights in population vulnerability by proposing a methodology for generating spatially referenced synthetic households. The methodology combines population synthesis techniques with population modelling and results in an output vector layer of households represented as points with unique structural and socioeconomic attributes. To demonstrate the effectiveness of the methodology, Chapter 4 provides a case study in Dumuria Upazila, Bangladesh, where an independent household survey is used to validate the location of the synthetic households and associated socioeconomic attributes.

Chapter 5 demonstrates how satellite remote sensing techniques to understand natural hazards can be integrated with the methodologies from previous chapters to target disaster response and recovery, as well as broader efforts to enhance resilience. The chapter focuses on an area of the Bangladesh Coastal Zone inhabited by 9.5 million people that was heavily affected by 2019 Cyclone Fani. The analysis presents new insights into the impact of the cyclone on populations by generating a high-resolution estimate of the flood extent, modelled point features of potentially flooded households, and modelled synthetic structural and socioeconomic characteristics of each household as a proxy for vulnerability.

Chapter 6 presents the conclusions reached from the research undertaken through this thesis. The chapter provides a summary of what has been achieved, including both theoretical and practical contributions for advancing global understanding of disaster risk. Chapter 6 concludes by presenting emerging areas of research in disaster risk management resulting from continuous improvements in the quality and quantity of raw data available paired with the ongoing development of novel and innovative modelling techniques.

2. The challenge of understanding disaster risk

2.1. Introduction

Investments in disaster risk management pay dividends over the lifetime of new assets; strengthening the resilience of infrastructure in developing countries alone presents a $4.2 trillion net benefit opportunity due to reduced economic impacts and fewer disruptions from hazard events (Hallegatte et al., 2019). However, without first having a firm understanding of risk in all its facets, it is impossible to make informed decisions. It is for this reason why the *Sendai Framework for Disaster Risk Reduction 2015-2030* (SFDRR) positioned its first priority for action around enhancing our understanding of disaster risk. This thesis focuses on the need to address the first priority for action of the SFDRR through the continued integration of increasingly higher-resolution data available through satellite Earth observation, localised data collection, and improved modelling techniques.

This chapter presents a review centred around the role that technological and methodological advancements have played in advancing our understanding of exposure, vulnerability, and hazard, and the applicability of those advancements towards informing disaster risk. First, the chapter presents the theoretical background of this thesis by outlining key definitions and concepts, including the three components of disaster risk (hazard, exposure, and vulnerability), the disaster risk management cycle (mitigation, preparedness, response, and recovery), and resilience. Next, the chapter provides background on the evolution of disaster risk reduction as a global priority, from the establishment of the United Nations Disaster Relief Office (UNDRO) in the 1970s to the *Hyogo Framework for Action 2005-2015* (HFA) and the SFDRR.

Lastly, this chapter focuses on recent advancements in satellite Earth observation, localised data collection, and modelling techniques that have provided new insights into our understanding of global disaster risk and provide the methodological foundation for this thesis. This includes a focus on three potential breakthroughs in risk information for developing countries pertinent to subsequent chapters: (i) mapping population exposure (ii) demonstrating more granular insights into population vulnerability; and, (iii) measuring the extent of flood events.

2.2. Definitions and key concepts

Disaster risk management is a rapidly evolving academic discipline. Accordingly, definitions and key concepts require regular revisiting, posing a challenge for researchers and practitioners alike. To address this challenge, the United Nations International Strategy for Disaster Reduction (UNISDR) has historically led the way by publishing incrementally updated terminology. In 2004, the UNISDR published standard terminology in *Living with risk: a global review of disaster risk reduction* (UNISDR, 2004). This was updated in 2009 with the *UNISDR terminology on disaster risk reduction* (UNISDR, 2009), and again in 2016 with the *Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction* (UN General Assembly, 2016).

2.2.1 Disaster risk as a function of hazard, exposure, and vulnerability

Disaster risk is defined as the "the potential loss of life, injury, or destroyed or damaged assets which could occur to a system, society or a community in a specific period of time" (UN General Assembly, 2016) and is often described as a function of hazard, exposure, and vulnerability; a concept known as the "Risk Triangle"

(Crichton, 1999). In the Risk Triangle, the area within the triangle represents the total risk, while each side represents one of the three factors that contribute to risk (Fig. 1).

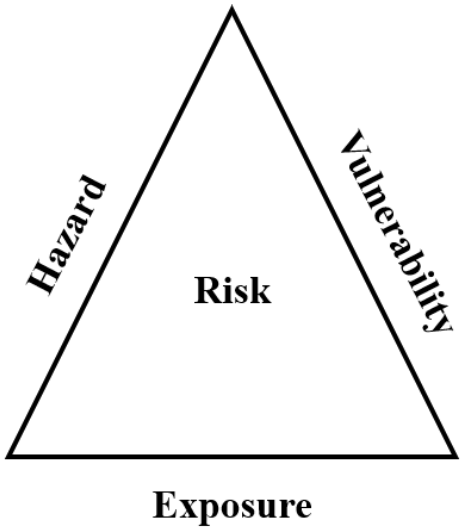


Fig. 1 The Risk Triangle (adapted from Crichton (1999))

A hazard is defined as a "process, phenomenon, or human activity that may cause loss of life, injury or other health impacts, property damage, social and economic disruption or environmental degradation." (UN General Assembly, 2016). Hazards can be either natural (resulting from natural Earth processes), biological, or man-made (Jingai et al., 1994; Gill and Malamud, 2014; Hewitt, 2014; Shi, 2019). This thesis focuses on natural hazards, which can be further categorised as either meteorological, climatological, hydrological, or geophysical (Bobrowsky, 2013; Munich Re, 2018). Because natural hazards are a result of natural Earth processes, they are inherently tied to spatial characteristics of the natural environment.

Exposure refers to "the situation of people, infrastructure, housing, production capacities and other tangible human assets located in hazard-prone areas" (UN General Assembly, 2016). Exposure can be further broken down into population

exposure and asset exposure (Garcia et al., 2016). Population exposure refers to the number of people located in a hazard-prone area, where damages from an event are reported as the number of people affected, or number of lives lost (Kulp and Strauss, 2017). Asset exposure typically refers to physical assets and their content (Amadio et al., 2019). Damages of physical assets may be reported as the number of houses, number of livestock, kilometres of roads, bridges, etc. and is reported in terms of its estimated monetary value (Jongman et al., 2012; Koks et al., 2015).

Lastly, vulnerability refers to "the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards." (UN General Assembly, 2016). Like exposure, vulnerability can be broken into population vulnerability and asset vulnerability. Population vulnerability focuses on the socioeconomic characteristics of exposed populations such as the level of poverty, access to health services, and demographic information (Chakraborty et al., 2005; Peduzzi et al., 2009), while asset vulnerability focuses on the physical characteristics of assets exposed such as the age, material, and number of floors in buildings, or classification sub-type of roads (e.g. rural road, county road, motorway) (Bell and Glade, 2004; Papathoma-Kohle et al., 2011).

Fig. 2 presents a simplified example of risk to population and risk to assets. The area in blue represents a modelled hazard (e.g. fluvial flooding). In total, there are five households containing a total of 12 people exposed to the hazard (population exposure), as well as five houses, one small business, and portions of five roads exposed to the hazard (asset exposure). Of the exposed populations, some may be more vulnerable than others to a hazard event (e.g. the elderly couple). Similarly, certain assets may be more vulnerable than others (e.g. rural road).

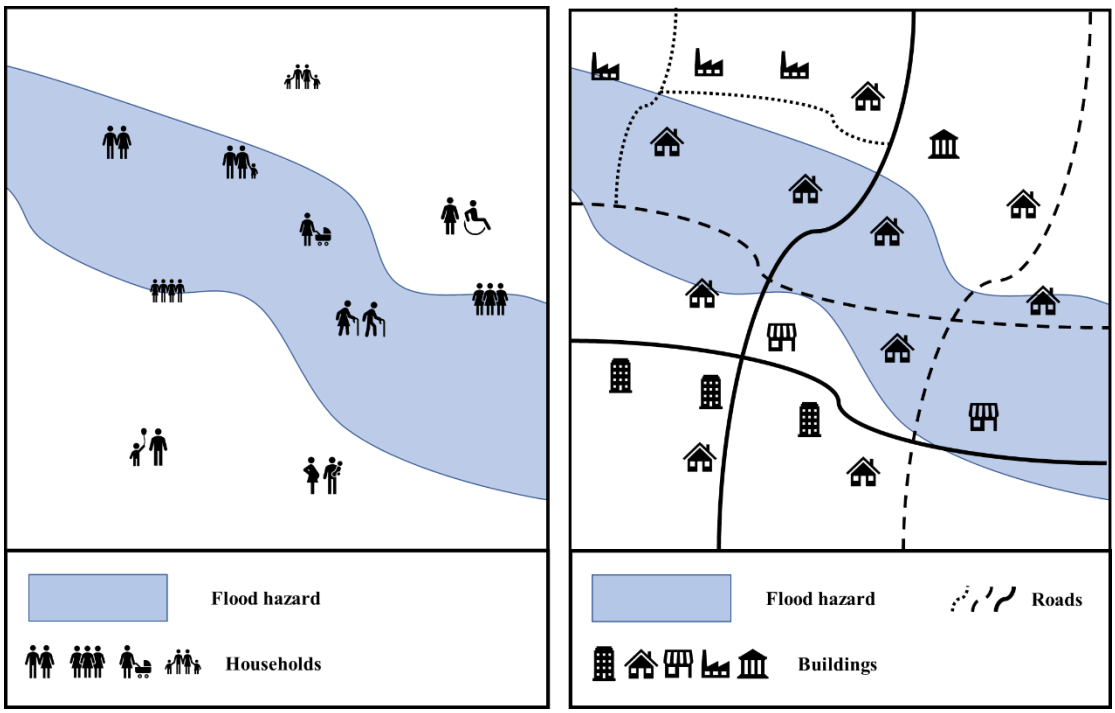


Fig. 2 Simplified example of risk to populations (left) and risk to assets (right)

2.2.2 Disaster risk management and resilience

The purpose of disaster risk management is to apply policies and strategies to prevent new disaster risk, reduce existing disaster risk, and manage residual risk (i.e. making the area of the Risk Triangle as small as possible) (UN General Assembly, 2016). Zoning policies that ban new construction in flood-prone areas are an example of measures that can be taken to prevent new disaster risk (Botzen et al., 2019). Disaster risk reduction efforts can focus on reducing the severity of a hazard (e.g. flood protective infrastructure), reducing the total population and assets exposed to a hazard (e.g. relocation of populations settled in landslide-prone areas), or reducing the vulnerability of populations exposed to a hazard (e.g. retrofitting buildings to be

seismic resilient). Lastly, while new disaster risk can be prevented and existing disaster risk can be reduced, there will remain some form of residual risk that requires managing. This may be due to the cost-benefit of undertaking measures for minor and infrequent events or related to the geographic extent of the hazard (e.g. countries such as Nepal, which sit entirely in an active fault zone). Residual risk can be managed through financial derivatives, reinsurance, and other forms of risk transfer (Kreimer et al., 2003).

Another key concept related to disaster risk is disaster resilience, which is defined as "the ability of a system, community or society exposed to hazards to resist, absorb, accommodate, adapt to, transform and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions through risk management" (UN General Assembly, 2016). Resilience-enhancing investments can help to avoid damages, but also reduce more wide-ranging consequences of a disaster, such as losses to livelihoods and welfare (Hallegatte et al., 2019). Natural disasters are responsible for approximately $520 billion of annual consumption losses and force 26 million people into poverty each year (Hallegatte et al., 2017). Focusing on a rapid restoration of services, livelihoods, and normalcy, are key to enhancing resilience, as losses may continue to mount for long after the hazard event has run its course. Disaster resilience can be improved through investments in disaster risk management which promote robustness, resourcefulness, redundancy, and rapidity in people, assets, and systems (Cimellaro et al., 2010).

Enhancing resilience requires prudent action across the four phases of the disaster risk management cycle: (i) mitigation; (ii) preparedness; (iii) response; and, (iv) recovery (Khan et al., 2008; Othman et al., 2014). Investments across the four

phases serve to reduce the damages and losses from a disaster event (Cimellaro et al., 2010). Mitigation and preparedness focus on preventing future disaster risks and minimising their effects, as well as preparing governments, communities, and individuals to better anticipate, respond to, and recover from a hazard event. Types of investments are diverse, and may include infrastructure works (e.g. river embankment, hospital), provision of goods (e.g. fire engines, emergency operations centres), risk financing instruments (e.g. insurance, reserve funds, contingent credit), systems planning (e.g. early warning, lifeline and critical infrastructure), capacity building, or the enforcement of policies (e.g. building codes) (Kelman, 2013). Collectively, investments in the mitigation and preparedness phases pay dividends by reducing the total damages and accumulated losses suffered from a disaster. The response phase focuses on the provision of public assistance and emergency services during and directly following a disaster event, with a primary intent to save lives, but also to ensure public safety, reduce health impacts, and provide subsistence for affected populations (UN General Assembly, 2016). In the recovery phase, investments shift focus towards the restoration of facilities and livelihoods. An emphasis on "building back better" can serve a double purpose by also mitigating future events (Hallegatte et al., 2018).

2.3. Disaster risk reduction as a global priority

Natural disasters are not a recent phenomenon; the remnants of the Roman city of Pompeii are still visible today, frozen in place as they were in 79 CE when an eruption of Mount Vesuvius covered the city under a thick carpet of volcanic ash that took the lives of approximately 2,000 people (Luongo et al., 2003). However, our

approach towards understanding disasters and our attempts to manage them have changed dramatically over time. From an initial focus on supporting response and recovery efforts, to a more recent embrace of preparation and mitigation, the global community has mobilised around disaster risk reduction as a critical priority for achieving sustainable development (Kelman, 2017).

2.3.1 From UNDRO to the Hyogo Framework for Action

Disaster risk management has been a priority of the United Nations since the 1960s, when the General Assembly began adopting measures regarding severe earthquakes in Iran and Yugoslavia, as well as a hurricane affecting much of the Caribbean. In the 1970s, the United Nations Disaster Relief Office (UNDRO) was formed to: (i) promote the study, prevention, control, and prediction of natural disasters; (ii) provide advice to governments on disaster planning; and, (iii) promote governments to improve national disaster warning systems. The UNDRO was further strengthened in the 1980s, where disasters in 1988 that included flooding in Sudan and Bangladesh, and a typhoon in the Philippines led to the General Assembly designating the 1990s as the "International Decade for Natural Disaster Reduction" and calling for the development of a framework for action. In 1994, a world conference was held in Yokohama, Japan, resulting in the Yokohama Strategy and Plan of Action for a Safer World, which became the first document providing international guidelines for disaster risk management (de la Poterie et al., 2015). The end of the decade saw the creation of the International Strategy for Disaster Reduction (ISDR) and the corresponding formation of a United Nations secretariat to support international collaboration around its objectives (UNISDR).

The 2000s saw a shift in global priorities away from response and recovery towards risk preparedness and mitigation emboldened in the Hyogo Framework for Action 2005-2015 (HFA) (Baudoin and Wolde-Georgis, 2015). The HFA put forward five priorities for action: (i) ensure that disaster risk reduction is a national and a local priority with a strong institutional basis for implementation; (ii) identify, assess and monitor disaster risks and enhance early warning; (iii) use knowledge, innovation and education to build a culture of safety and resilience at all levels; (iv) reduce the underlying risk factors; and, (v) strengthen disaster preparedness for effective response at all levels (UNISDR, 2005). The adoption of the framework was complemented by the establishment of different mechanisms designed to help implement disaster risk reduction activities at national and local levels in addition to working transnationally. The Global Facility for Disaster Reduction and Recovery (GFDRR) was subsequently established in 2006 as a global partnership that helps developing countries better understand and reduce their vulnerability to natural hazards and climate change. The following years saw the widespread adoption of legislation on disaster management, the formation or strengthening of disaster management agencies, and a broader focus on mainstreaming disaster risk reduction, particularly across countries in the developing world (Olowu, 2010; de la Poterie et al., 2015).

2.3.2 Priority 1 of the Sendai Framework for Disaster Risk Reduction: understanding risk

The HFA was a ten-year plan, effective only until 2015. Accordingly, the SFDRR was adopted to continue the progress made, renew commitments, and reassess priorities

for the subsequent 15 years. The SFDRR aims to achieve "the substantial reduction of disaster risk and losses in lives, livelihoods and health and in the economic, physical, social, cultural and environmental assets of persons, businesses, communities and countries" (UN General Assembly, 2015). Compounding challenges such as climate change are reflected in the SFDRR, as well as the important role that the adoption of new technologies can serve to enhance our knowledge of risks and contribute towards early warning systems (de la Poterie et al., 2015; Kelman, 2015). The SFDRR put forward four priorities for action: (i) understanding disaster risk; (ii) strengthening disaster risk governance to manage disaster risk; (iii) investing in disaster risk reduction for resilience; and, (iv) enhancing disaster preparedness for effective response and to "Build Back Better" in recovery, rehabilitation, and reconstruction.

The first priority for action of the SFDRR focuses on understanding disaster risk in all its dimensions, including hazard, exposure (populations and assets), and vulnerability, due to its importance for informing investments across the disaster risk management cycle. Robust risk information is critical for pre-disaster risk assessments, mitigation planning, the development of appropriate preparedness measures, and the coordination of response efforts following a disaster event. Each priority for action in the SFDRR proposes key activities to guide stakeholders at national and local levels, as well as regional and global levels. Table 2 presents the key activities listed for achieving the first priority for action, understanding disaster risk, at national and local levels.

Table 2. *Key activities listed in the SFDRR for understanding disaster risk at national and local levels*

Activity	***Priority 1: Understanding disaster risk***
1.	To promote the collection, analysis, management and use of relevant data and practical information and ensure its dissemination, taking into account the needs of different categories of users, as appropriate
2.	To encourage the use of and strengthening of baselines and periodically assess disaster risks, vulnerability, capacity, exposure, hazard characteristics and their possible sequential effects at the relevant social and spatial scale on ecosystems, in line with national circumstances
3.	To develop, periodically update and disseminate, as appropriate, location-based disaster risk information, including risk maps, to decision makers, the general public and communities at risk of exposure to disaster in an appropriate format by using, as applicable, geospatial information technology
4.	To systematically evaluate, record, share and publicly account for disaster losses and understand the economic, social, health, education, environmental and cultural heritage impacts, as appropriate, in the context of event-specific hazard-exposure and vulnerability information;
5.	To make non-sensitive hazard-exposure, vulnerability, risk, disaster and loss-disaggregated information freely available and accessible, as appropriate
6.	To promote real time access to reliable data, make use of space and in situ information, including geographic information systems (GIS), and use information and communications technology innovations to enhance measurement tools and the collection, analysis and dissemination of data
7.	To build the knowledge of government officials at all levels, civil society, communities and volunteers, as well as the private sector, through sharing experiences, lessons learned, good practices and training and education on disaster risk reduction, including the use of existing training and education mechanisms and peer learning
8.	To promote and improve dialogue and cooperation among scientific and technological communities, other relevant stakeholders and policymakers in order to facilitate a science policy interface for effective decision-making in disaster risk management
9.	To ensure the use of traditional, indigenous and local knowledge and practices, as appropriate, to complement scientific knowledge in disaster risk assessment and the development and implementation of policies, strategies, plans and programmes of specific sectors, with a cross-sectoral approach, which should be tailored to localities and to the context
10.	To strengthen technical and scientific capacity to capitalise on and consolidate existing knowledge and to develop and apply methodologies and models to assess disaster risks, vulnerabilities and exposure to all hazards
11.	To promote investments in innovation and technology development in long-term, multi-hazard and solution-driven research in disaster risk management to address gaps, obstacles, interdependencies and social, economic, educational and environmental challenges and disaster risks
12.	To promote the incorporation of disaster risk knowledge, including disaster prevention, mitigation, preparedness, response, recovery and rehabilitation, in formal and non-formal education, as well as in civic education at all levels, as well as in professional education and training
13.	To promote national strategies to strengthen public education and awareness in disaster risk reduction, including disaster risk information and knowledge, through campaigns, social media and community mobilisation, taking into account specific audiences and their needs
14.	To apply risk information in all its dimensions of vulnerability, capacity and exposure of persons, communities, countries and assets, as well as hazard characteristics, to develop and implement disaster risk reduction policies
15.	To enhance collaboration among people at the local level to disseminate disaster risk information through the involvement of community-based organisations and nongovernmental organisations

Source: UN General Assembly, (2015)

Five years following the adoption of the SFDRR, large challenges remain for achieving national and local level progress towards the four priorities for action. Progress towards understanding risk has proven to be particularly challenging (Mizutori, 2020). Noted gaps at the national and local levels include, *inter alia*: (i) the need for improved data collection, analysis, storage in public access databases, local capacity building, and enabling policy guidelines for information management; (ii) the need for more free and open high-resolution remote sensing data from satellites to supplement local observations; and, (iii) the need for open access to advanced software packages which can help agencies derive more insights from geospatial data (Rahman and Fang, 2019). These challenges are further complicated by continuous technological advancements which require monitoring and a high level of capacity to discern their potential application for informing a better understanding of disaster risk.

2.4. Towards a high-resolution understanding of global disaster risk

In the years since the adoption of the SFDRR, advancements in Earth observation, extended data collection methods, and modelling have enabled increasingly detailed and granular insights into our understanding of global disaster risk. The growing availability of commercial satellite data through philanthropy partnerships paired with the recent launches of Sentinel 1 and 2 satellites have led to higher-resolution Earth observation data becoming available for understanding the built and natural environments at a much larger volume and scale in developing countries (Paganini et al., 2018). Localised data providing further insights into attributes of physical assets and the socioeconomic characteristics of populations are also becoming more readily

available. This has been facilitated by increasing rates of mobile data penetration in the developing world, global movements focused on collecting and promoting open data, and advancements in data collection techniques and accompanying sensors (Bahia and Suardi, 2019). When paired with novel forms of analysis, particularly those that rely on machine learning techniques, the vast amounts of newly available data can be applied to improve existing data sets and unlock novel derived data sets with vast potential to inform global disaster risk.

2.4.1 Advancements in global Earth observation

In 1972, Landsat 1 was launched by the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS) as the first non-military Earth observation satellite. In 1986, SPOT-1 was launched as the first commercially purposed Earth observation satellite, offering 10-metre panchromatic (black and white) and 20-metre multispectral (visible green, visible red, near infrared, etc.) resolution capability. Further breakthroughs occurred following the launch of IKONOS in September 1999, which became the first commercially available high-resolution Earth observation satellite. IKONOS offered panchromatic and multispectral imagery at 1- and 4-metre resolution with a revisit time of approximately three days until its retirement in 2015. In the subsequent years, a variety of increasingly high-resolution commercial Earth observation platforms have emerged, generating larger constellations that provide for more frequent revisit times, and including more sophisticated sensors.

Another set of breakthroughs occurred following a decision by the United States Department of Commerce in 2014 to allow for DigitalGlobe to provide imagery

for commercial purposes from its constellation (notably Worldview-3/4 satellites) at up to 25-centimetre resolution for panchromatic and 100-centimetre resolution for multispectral ground sample distance (GSD). Airbus followed suit and currently provides competition through its own constellation (notably Pleiades 1a/1b and SPOT 6/7). Synthetic aperture radar (SAR) sensors (e.g. Cosmo-SkyMed, TerraSAR-X) have enabled collection of high-resolution data independent of weather conditions, while multi-view and adjustable sensors have enabled preferential observation as well as three-dimensional renderings of the natural and built environment (Huang et al., 2018; Wohlfart et al., 2018). Planet Labs deploys much larger constellations (more than 130 as of May 2020) of much smaller satellites ("Doves" each weigh only 4 kilograms), offering daily revisit times for most of the Earth's surface at 3-5-metre optical resolution. Twice-daily revisit times for targeted locations are currently available at 72-centimetre resolution through Planet Labs, while 30-centimetre resolution will be available through Airbus with the launch of Pleiades Neo in 2020, providing more frequent high-resolution opportunities to monitor changes and pinpoint trends.

Despite rapidly lowering prices for high-resolution commercial Earth observation data and the rise of data philanthropy (e.g. DigitalGlobe Foundation), cost remains a binding constraint for large-scale applications, particularly in many developing countries (Paganini et al., 2018). Fortunately, recent years have also led to the proliferation of freely open satellite Earth observation data supplied by governments progressing global development solutions (Anderson et al., 2017). In 2007, the USGS began by making Landsat 7 data available at no cost, and in 2014, the European Space Agency (ESA) adopted a similar approach for Sentinel 1-A and all further satellite missions as part of the Copernicus Programme. The advantages of

opening non-sensitive data have become clear: only for the year 2011, the economic benefit of making Landsat data open was calculated to be $2.1 billion (Zhu et al., 2019). Landsat provides over 45 years of data available in archive for free public usage, with Landsat 7/8 currently capturing multispectral data at 30-metre (15-metre enhanced) spatial resolution and a revisit time of up to eight days. The Copernicus Programme comprises of a series of Sentinel satellites, which provide free and open SAR data at up to 5-metre resolution (Sentinel-1 A/B, launched in 2014 and 2016) and multispectral data at up to 10-metre resolution (Sentinel-2 A/B, launched in 2015 and 2017) with a goal of eventually supporting revisit coverage of once every two days (Torres et al., 2012). Large volumes of open Earth observation data are increasingly available on a variety of cloud-based platforms that enable broader access. For example, Google Earth Engine, launched in 2010, enables users to visualise and analyse data from both Landsat and Sentinel satellites through Google Cloud Storage.

2.4.2 Advancements in localised data collection

Between 2014 and 2018, mobile internet connectivity in Sub-Saharan Africa increased from 13 percent to 24 percent, while smartphone penetration trebled from 10 percent to 30 percent (Bahia and Suardi, 2019). Between 2016 and 2018, the average monthly cost of mobile internet data halved in the region as well, rendering services more affordable for millions. Smartphones can serve a variety of purposes, including the collection of location-based data that can complement and improve upon Earth observation data while providing a means for collection of non-tangible socioeconomic attributes and consumer data critical for understanding population

vulnerability. Smartphones measure location precisely using a combination of Wi-Fi positioning systems, multilateration between cell towers and the phone, and global navigation satellite systems (e.g. the Global Positioning System) (Toch et al., 2019). Merry and Bettinger (2019) measure horizontal position accuracy of smartphones at 7-13 metres in an urban environment (consistent with recreational-grade GPS receivers) while Tomastik et al. (2016) demonstrate that less than 2 metres is possible in an open area.

Applications such as OpenDataKit (ODK) have been developed as open-source tools to facilitate the collection of geospatially referenced data through smartphones in developing countries (Hartung, 2010). The utility of ODK was exemplified during the 2014-2016 West Africa Ebola outbreak to improve the timeliness of reporting and provide accountability in contact tracing (Tom-Aba et al., 2015). Following the 2015 Earthquake in Nepal, tablets with mobile internet data were used to collect georeferenced information regarding the structural damage to over one million houses for the purposes of selecting beneficiary households to receive government grants for "build back better" owner-driven reconstruction (Lallemant at al., 2017). Georeferenced household surveys have also been applied after disaster events to better understand the role of poverty in the impact of floods on households (Erman et al., 2019). Recent years have seen an increase in the large-scale analysis of human mobility data including tracking of population movements and better targeting of populations in need of response and recovery efforts following a disaster event (Bengtsson et al., 2011; Wilson et al., 2016; Maas et al., 2019).

The rise of smartphones has also corresponded with a rise in crowdsourced data and collaborative mapping efforts. OpenStreetMap (OSM), for example, has over 6 million registered contributors who have mapped over 5 billion nodes and 500

million ways as of May 2020. Contributors can create new features or edit existing ones using online or smartphone-based software tools and a basemap of satellite imagery. Importantly, OSM can tag attributes to features pertaining to their physical characteristics, such as the number of storeys in a building or the building type (fire station, school, residential, etc.), critical for providing insights into asset and population vulnerability that may be difficult to detect at scale using only remote sensing. OSM data has demonstrated its value for disaster risk management in data scare environments where government data may be less robust (Poiani et al., 2016; Goldblatt et al., 2020). Another crowdsourced platform, Mapillary, provides geotagged photos which can link to OSM edits and provide additional granularity on the physical vulnerability of assets (Juhasz et al., 2016). Efforts to detect more detailed physical attributes of the built environment have been aided by lowering costs for collecting and processing street view imagery and localised remote sensing data (e.g. very high-resolution imagery and lidar) through advancements in sensor technology and collection methods (e.g. unmanned aerial vehicles (UAVs) (Liu et al., 2017) and backpack-mounted street view cameras) (Nocerino et al., 2019).

Less technology driven yet still important, continuous advancements in data collection and stewardship have helped to shed more light on the socioeconomic characteristics of populations. Initiatives such as the Integrated Public Use Microdata Series (IPUMS) have worked to populate a database of individual-level microdata samples, including from international census records. Efforts such as the Demographic and Health Survey (DHS) program, the Living Standards Measurement Study (LSMS), and the Multiple Indicator Cluster Surveys (MICS) have also helped to provide baseline socioeconomic information in developing countries. Although inherent privacy concerns surrounding socioeconomic data limit the volume and

spatial granularity of microdata releases, the quality has improved dramatically with advancements in anonymisation techniques (Cassa et al., 2008) and broader support from governments for open data initiatives (Attard et al., 2015).

2.4.3 Advancements in modelling techniques

Modelling and analysis techniques have advanced dramatically in parallel to the improvements in data collection that have provided for larger volumes of higher-resolution and more granular data since the adoption of the SFDRR. Catastrophe models, which aim to quantify the financial impact of future catastrophic events, have long been used by the (re-) insurance industry to assess and manage losses from disaster events. Following Hurricane Andrew, which struck Florida in 1992, the private firm Applied Insurance Research (AIR) estimated insured property losses of $13 billion within four hours of landfall using a catastrophe model, much higher than the $7 billion estimated elsewhere and much closer to the $15.5 billion eventually realised (Clark, 2002). As a result of the damage caused by Hurricane Andrew, 13 insurers became insolvent and improving catastrophe modelling became a key focus for the industry in the subsequent years.

Catastrophe modelling is a sub-set of disaster risk modelling largely reserved for the (re-)insurance industry that focuses on generating information such as: (i) exceedance probability (EP) curve, which refers to the annual probability a certain loss threshold is exceeded; (ii) probable maximum loss (PML); and, (ii) average annual loss (AAL), which represents the minimum premium required to cover modelled losses. A public decision-maker, for example, may also interested in other facets of risk, such as the number of people affected or damage to lifeline and critical

infrastructure in a specific scenario (Bostrom et al., 2008). There are two types of disaster risk models: probabilistic and deterministic (UNISDR, 2015a). Probabilistic models reflect all possible scenarios, their probability, and accompanying impacts, while deterministic models only consider the impact of a single scenario. To accurately model damages and losses, high-quality historical records of hazard events, exposure information (e.g. replacement value, occupancy, etc.), and vulnerability of assets and populations are essential (Linnerooth-Bayer and Hochrainer-Stigler, 2015). Appropriate data can be difficult to attain in developing countries, often limiting the ability of insurance markets to expand and reducing the utility of disaster risk assessments for public decision-making (Grossi and Windeler, 2005).

The past decades have seen more types of hazards modelled (e.g. seismic, flood, wind speed, heat, landslide) using increasingly advanced methodologies paired with higher quality input data. This has led to the proliferation of multi-hazard risk models (El Morjani et al., 2007; Kappes et al., 2012). At a global scale, models such as Hazus (Schneider and Schauer, 2006), RiskScape (King and Bell, 2006), and CAPRA (Cardona et al., 2010) all offer physically based, deterministic methods for multiple hazards (e.g. Hazus offers earthquakes, floods, hurricanes, and tsunamis). Global flood modelling in particular has seen significant advancements resulting from new input data sets derived from satellite Earth observation including digital elevation models (Rodriguez et al., 2006), river networks (Lehner et al., 2008), and river channel width (Yamazaki et al., 2014). Better input data combined with new modelling methods and improvements in algorithms have enabled up to 100-metre resolution gridded flood risk data to be generated from global computational models (Trigg et al., 2016). Some examples of global gridded flood risk data include

GLOFRIS (Winsemius et al., 2013), JRC (Dottori et al., 2016), and SSBN (Sampson et al., 2015).

Advancements in machine learning (a subset of artificial intelligence) have played a particularly significant role in enabling the multitude of newly derived spatial data sets available at a global scale that can serve as inputs for various disaster risk models. Traditional geostatistical techniques are limited in their application due to certain peculiarities of geospatial phenomena including nonlinearity, multi-scale variability, non-stationarity (spatial and temporal), and the presence of noise and outliers (Kanevski et al., 2008). Machine learning is an approach to achieve artificial intelligence by applying algorithms to parse data, learn from it, and make predictions (Alom et al., 2018). In machine learning, one must "train" the machine to complete a task through large volumes of data and algorithms.

Deep learning represents an advancement in machine learning that relies on artificial neural networks, which are inspired by our understanding of the brain. Deep learning is computationally intensive, and it was not until the widespread adoption of graphics processing units (GPUs) paired with advancements in algorithms that deep learning efforts were able to propagate (Schmidhuber, 2015). One area of deep learning that has proven its utility for geospatial data is computer vision, which trains computers, in effect, to see (Demir et al., 2018). Computer vision techniques include, *inter alia*, image classification (Maggiori et al., 2016; Ma et al., 2017), object detection (Cheng and Han, 2016; Sevo and Avramovic, 2016), semantic segmentation (Li et al., 2017), and instance segmentation (Liu et al., 2018). Another area where deep learning has proven its utility is in super-resolution and image fusion for statistical downscaling and blending, which serves to enhance the resolution of baseline data (Belgiu and Stein, 2019; Reichstein et al., 2019).

Deep learning has become relevant for understanding risk in all its facets. For hazard data, deep learning has shown a range of utility, including predicting flash flood susceptibility (Bui et al., 2020), improving the resolution of global coastal digital elevation models for flood modelling (Kulp and Strauss, 2018), understanding the patterns of aftershocks from large earthquakes (DeVries et al., 2018), and landslide detection (Ghorbanzadel et al., 2019). Deep learning has also provided novel insights into the built and natural environment relevant to exposure mapping, including vehicle instance segmentation (Mou and Zhu, 2018), semantic segmentation of slums in satellite images (Wurm et al., 2019), segmentation of building footprints from aerial and satellite imagery (Ji et al., 2018; Li et al., 2019), and land-use changes from convolutional neural networks (Zhao and Du, 2016). When combined with locally collected socioeconomic data (e.g. from surveys) and satellite Earth observation data, machine learning techniques can also help to predict indicators of social vulnerability such as poverty (Jean et al., 2016; Xie et al., 2016).

At a global level, many of the providers of commercial satellite Earth observation data such as DigitalGlobe and Planet have developed business lines around analytical products and derived data sets generated through the application of machine learning techniques. The resolution and accuracy of global data sets which directly classify features of the natural and built environment (e.g. land cover, built-up area, etc.), as well as those which are derived using several spatial covariates (e.g. population density) have benefited as a result. Locally collected and crowdsourced data can also be applied towards improving global data products by providing a means for necessary training data of machine learning models. For example, OpenStreetMap data have been used as training data for rapid land use and land cover classification of Landsat satellite imagery (Johnson and Iizuka, 2016). Training data

are particularly critical for adding granularity to a global understanding of asset and population vulnerability, the proxies of which (e.g. building type classification and poverty) tend to be more difficult to detect through remote observation but can be correlated to certain detectable patterns or features of the built environment (Gram-Hansen et al., 2019). Novel applications of statistical regression-based and spatial microsimulation techniques for small-area estimation (SAE) that pair aggregate releases with microdata releases can also help to inform socioeconomic characteristics of large populations at a higher resolution. Spatial microsimulation provides an emerging bottom-up approach to downscaling, where representative synthetic populations are generated for each target area by "fitting" microdata to multidimensional aggregate tables (Whitworth et al., 2017).

2.5. Key breakthroughs in disaster risk information for developing countries

Larger volumes of higher resolution satellite Earth observation data, improvements to localised data collection, and broad trends pushing governments and firms towards embracing open and crowdsourced data have led to a continuous proliferation of ever-improving sources of spatial data. When paired with emerging techniques for statistical modelling and machine learning, further insights from derived data can be drawn upon for informing hazard, exposure, and vulnerability. These developments have enabled continuous advancements in our understanding of risk at scale in developing countries, including through their integration into physically based risk models important for adaptation planning. Specifically, three breakthroughs deserve closer examination given their potential for disrupting traditional methods of

informing disaster risk: (i) rapid improvements to mapping population exposure resulting from coupling dasymetric modelling techniques with newly derived spatial layers of the built environment; (ii) novel approaches for demonstrating spatial granularity of population vulnerability by generating representative synthetic populations and pairing them with high-resolution population exposure data; and, (iii) quicker, more accurate, and higher resolution flood event detection enabled by synthetic aperture radar (SAR) data becoming openly available through Sentinel-1.

2.5.1 Mapping population exposure

Several recent efforts applying advanced modelling techniques have produced global-continental- or country-extent gridded population data (counted at place of domicile) at increasingly high levels of resolutions (Leyk et al., 2019). Openly available country-extent gridded population data is now offered at up to 30-metre resolution for the year 2015 through the High-Resolution Settlement Layer (HRSL), and annually at 100-metre resolution from 2000 until 2020 through WorldPop. Recent efforts utilising the World Settlement Footprint (WSF) suite of products have tested methods resulting in up to 10-metre resolution for the year 2015 (Palacios-Lopez et al., 2019). At the time of the SFDRR's adoption in 2015, prevalent open access country-extent gridded population data were limited to 1-kilometre resolution (Global Rural Urban Mapping Project (GRUMP) v1, and Global Population of the World (GPW) v4) with the exception of early efforts by the team that would later become WorldPop. Table 3 provides a timeline of advancements in resolution of prominent gridded population data sets that cover multiple countries.

Table 3. *Prominent open access gridded population data sets that cover multiple countries by first year of release*

First year of release	Data set	Resol-ution	Method	Source	Reference Link
1995	GPW v1	10 km	Pycno-phylactic	Tobler et al., 1997	sedac.ciesin.columbia.edu/data/collection/gpw-v1
1998*	LandScan	1 km	Smart interpolation	Dobson et al., 2000	landscan.ornl.gov
2000	GPW v2	5 km	Areal weighting	Deichmann et al., 2001	sedac.ciesin.columbia.edu/data/collection/gpw-v2
2004	GRUMP v1	1 km	Dasymetric	Balk et al., 2005	sedac.ciesin.columbia.edu/data/collection/grump-v1
2013**	WorldPop	100 m	Statistical/ dasymetric	Stevens et al., 2015	www.worldpop.org
2015***	GPW v4	1 km	Areal weighting	Doxsey-Whitfield et al., 2015	sedac.ciesin.columbia.edu/data/collection/gpw-v4
2016	GHS-POP	250 m	Dasymetric refinement	Freire et al., 2016	ghsl.jrc.ec.europa.eu/ghs_pop2019.php
2016	HRSL	30 m	Binary dasymetric	Tiecke et al., 2017	www.ciesin.columbia.edu/data/hrsl
2019	WSF	10 m	Binary dasymetric	Palacios-Lopez et al., 2019	N/A

* LandScan is open access for research but not for commercial usage

**WorldPop united AfriPop, AsiaPop, and AmeriPop projects which had already released 100-metre resolution data sets prior to 2013 (Linard et al. 2012)

***Have not included GPWv3 as it did not improve the resolution of release or significantly change upon the methodology from GPWv2

Source: Balk et al. (2005) and Leyk et al. (2019)

Improvements in gridded population modelling have been driven by several factors, including: (i) greater availability of robust census data at finer administrative levels (e.g. district level vs. sub-district level); (ii) advancements in modelling techniques; and, (iii) greater selection of spatial covariate data sets at higher levels of resolution for dasymetric approaches. While it is possible to estimate populations counts by employing bottom-up, survey-driven approaches (Weber et al., 2018) or methods using mobile phone data (Deville et al., 2014), most high-resolution techniques utilise population censuses where available as both a baseline for spatial disaggregation, and as a measure to calculate the accuracy of the technique employed (Wardrop et al., 2018). Thus, gaining wider access to existing higher resolution

census data and providing countries with the tools to conduct more robust censuses in the future remains a top priority.

Areal interpolation refers to one category of methods to spatially disaggregate populations where geographic information from one source area is transformed to another target area (Tobler et al., 1997; Liverman et al., 1998, Mennis, 2003; Wu et al., 2005). Methods such as areal weighting or inverse distance weighted interpolation assume that the population surface is distributed homogeneously within source areas, meaning that the probability that a person lives with a certain location is always greater than zero (Nagle et al., 2014). Dasymetric modelling is a subtype of areal interpolation that instead approaches the population surface as discrete with sharp boundaries identified using ancillary variables (Mennis, 2003).

Modelling techniques for generating gridded population data over large extents have evolved substantially over time. GPW v1 distributed populations proportionally within an administrative boundary, while applying a smooth pycnophylactic interpolation redistribution algorithm on grid cells that crossed one or more administrative boundaries (Balk et al., 2006). GPW v2, GPW v3, and GPW v4 all relied on a simpler form of areal weighting for grid cells that crossed one or more administrative boundaries. Improvements in accuracy across subsequent versions can be credited to drastic increases in the total number of administrative units considered (19,000 for GPW v1, 120,000 for GPW v2, 375,000 for GPW v3, and 12,500,000 for GPW v4) (Balk et al., 2006; Doxsey-Whitfield et al., 2015). In 1998, LandScan introduced a smart interpolation approach for generating global gridded population data that relied on probability coefficients tied to road proximity, slope, land cover, and night-time lights (Dobson et al., 2000). The LandScan methodology was the first to rely on satellite data as key inputs to a global gridded population data set and the

first to offer 1-kilometre resolution with a focus on ambient populations. In 2004, GRUMP v1 adapted the GPW approach of areal weighting to offer a 1-kilometre resolution global gridded population data set by explicitly considering population of urban areas. The methodology segregated urban footprints (often from night-time lights) and incorporated population estimates outside of statistical reporting units (Balk et al., 2006). GRUMP v1 was the first global gridded population data set to employ basic dasymetric modelling methods.

The opening of Landsat and Sentinel data paired with novel machine learning methods for extensive object detection led to further advancements in population modelling in the mid-2010s. In particular, the increasing availability of higher resolution and more granular derived data of the natural and built environment provided a better baseline of potential spatial covariates for dasymetric modelling techniques. The Global Human Settlement Population Grid (GHS-POP) utilises a dasymetric refinement approach reliant on built-up areas from the GHS-BUILT layer (38-metre resolution derived from Landsat satellite data) to generate a 250-metre resolution global gridded population data set (Freire et al., 2016). WorldPop generates 100-metre resolution country-extent gridded population data by employing a weighting scheme based on a random forest, nonparametric predictive model which improves incrementally as existing spatial covariates are upgraded and more are included (Stevens et al., 2015). Accordingly, WorldPop continues to refine its approach to include the best possible data as it becomes available. HRSL employs a much simpler binary dasymetric approach to produce country-extend gridded population data at 30-metre resolution, where the only spatial covariate considered by the model is building footprints detected using convoluted neural networks from satellite imagery (Tiecke et al., 2017). A similar approach applies WSF layers for

extent of human settlement (WSF-2015) and density of impervious surfaces (WSF-2015-Density) derived from Landsat-8 and Sentinel-1 to generate 10-metre resolution country-extend gridded population data (Palacios-Lopez et al., 2019).

Having a precise and accurate baseline layer of population exposure is critical for understanding risk. The improvements in the resolution of global gridded population data since the adoption of the SFDRR in 2015 have been substantial, particularly in developing countries. However, questions remain regarding the utility of global gridded population data released at such high resolutions, particularly in countries with poor baseline census data (outdated, incomplete, or unavailable at finer administrative levels) and in areas with large local variations in population density (e.g. cities) (Linard and Grippa, 2019). This represents a particularly acute challenge for top-down methods which heavily rely on a strong binary masking variable (e.g. HRSL). For example, building footprints or built-up areas do not consider the average floor space per person (thus potentially underreporting populations in slums), the height of buildings, or the building type classification (e.g. commercial, residential, industrial). Local efforts have already incorporated these types of supplementary attributes into population layers (Sridharan and Qiu, 2013), while global efforts are emerging as well (Helber et al., 2019; Palacios-Lopez et al., 2019). In parallel to ongoing improvements to existing methods, more needs to be done to understand the opportunities and constraints of applying different global- and country-extent gridded population data for decision-making around disaster risk management.

2.5.2 Demonstrating more granular insights into population vulnerability

The significant advancements in mapping global population exposure at increasingly higher resolutions have yet to lead to similar advancements in mapping population vulnerability (i.e. the welfare of populations) at scale. When a hazard event occurs, poor households are more frequently exposed, lose a greater share of their wealth ($1 in losses means something much different to a poor household than it does to a rich household), and receive less financial support to recover (Hallegatte et al., 2017). Thus, a granular understanding of the socioeconomic characteristics of households affected by a hazard event and the extent to which they are affected are both important for ensuring effective disaster risk management and poverty reduction efforts. At 10-, 30-, or 100-metre resolution, gridded population exposure models often allocate only a couple of people or households per populated grid cell, particularly in more remote areas. However, without any sort of socioeconomic characteristics associated to those people or households, understanding their relative vulnerability to a hazard event remains a challenge. In developing countries where data relevant to understanding poverty at a granular level are often less available, the challenge can be particularly acute (Serajuddin et al., 2015).

At a localised level, there has been substantial progress to understand the spatial dynamics of population vulnerability. Certain characteristics of the built environment detectable by Earth observation such as small roof sizes, high density, narrow streets, or heterogeneous building orientation, have been used to characterise slum areas which tend to be inhabited by more vulnerable populations (Kuffer et al., 2016). However, these characteristics may differ dramatically depending on the city

or country context, making any efforts to scale methodologies across geographies challenging (Kit et al., 2012; Kuffer et al., 2016). Slums and deprived areas are more often undercounted in official censuses and may not appear on authorised maps if informal or illegal (Kuffer et al., 2016). Although mapping these areas presents an opportunity to develop pro-poor policies and protect resident rights, it also presents a tangible privacy concern in the form of mass evictions that must be considered alongside. This is an acute challenge in disaster risk management, where poor populations may experience an exposure bias by being pushed towards hazard-prone areas (Baker, 2012; Winsemius et al., 2015; Alvarez and Cardenas, 2019). Seismic vulnerability of physical structures has also been mapped at localised scales using building characteristics as inputs (Costanzo et al., 2016). In the 2015 Nepal Earthquake, most damaged houses were constructed with stone or brick masonry, while most reinforced concrete houses remained undamaged (Goda et al., 2015). The cost of reinforced concrete houses tends to be much higher, meaning that building type relates to both asset vulnerability and population vulnerability. Although certain characteristics of the natural and built environment detectable at scale from satellite Earth observation may serve as proxies for population vulnerability, often the relationships are much more complex (Jean et al., 2016).

Poverty is typically measured through consumption-based welfare indicators or through multidimensional indices (Ravallion, 2011). Consumption-based welfare models rely on household surveys that provide reasonable measures of consumption (Elbers et al., 2003), while multidimensional indices rely on additional factors, such as those related to education, health, and standard of living (Alkire and Santos, 2014). Efforts to map poverty at a subnational scale often rely on small-area estimation (SAE) methods that apply econometric techniques to downscale socioeconomic data.

Aggregate releases of data (typically census data) are paired with microdata from the same data set or a separate survey (e.g. the Demographic and Health Survey (DHS) or the Living Standards Measurement Survey (LSMS)), which provide matching individual or household attributes for a percentage of the population within the aggregated geographic unit.

Although many unique methodologies have been proposed for SAE (e.g. Bishop et al., 2007; Ghosh and Rao, 1994; Whitworth, 2013), they typically fall under one of two categories: statistical regression-based techniques or spatial microsimulation techniques. Statistical SAE techniques, such as the ELL method (Elbers et al., 2003), match explanatory variables from the higher-level data set with those available from a separate data set in the smaller area of interest. This has proven a particularly robust technique for downscaling aggregates (e.g. administrative level 2 to administrative level 4) since results are inherently linked to confidence intervals around central point estimates. Statistical SAE techniques therefore concretely allow for uncertainty to be considered, and for downscaling to stop at a level where the analyst determines it is no longer of value. Although advancements have been made to calculate uncertainty using certain spatial microsimulation techniques, it largely remains a methodological weakness of the technique when used for SAE (Whitworth et al., 2017). The primary advantage of using a spatial microsimulation technique is that it results in the generation of a synthetic population for each target area. These synthetic populations can be utilised in distinct forms of modelling, including bottom-up approaches for understanding complex geographic systems. Synthetic populations have been used as inputs to agent-based simulation models for the transport sector (Voellmy et al., 2001), coupled socio-ecological systems (Filatova et al., 2013), and health (Tracy et al., 2018).

Pairing methods which downscale socioeconomic data with Earth observation data offers an emerging opportunity to provide further granularity for understanding population vulnerability at scale. For example, a simple linear regression model using a variety of inputs such as density of buildings, number of cars, and roof material, was able to explain nearly 60 percent of poverty headcount rates and average log consumption in Sri Lanka (Engstrom et al., 2017). A variety of innovative methods which pair spatial covariates derived from Earth observation data with advanced modelling techniques have been applied towards poverty mapping efforts in least-developed countries (Ghosh et al., 2013; Tatem et al., 2014; Jean et al., 2016; Xie et al., 2016; Steele et al., 2017). With higher resolution gridded population exposure data now available for most countries globally, advancements in Earth observation data, and improved methods for assessing multidimensional poverty, more can be done to better understand the vulnerability of populations at a granular level.

2.5.3 Measuring the extent of flood hazard events

A granular understanding of population exposure and vulnerability prior to a hazard event can help to quickly target response and recovery efforts critical for saving lives and reducing longer term welfare losses. However, this only holds true if one also has an accurate and timely understanding of the area affected. Between 1995 and 2015, floods accounted for 43 percent of all documented natural disasters, affecting 2.3 billion people and causing $662 billion in damage (UNISDR, 2015b). Until recently, efforts to quickly assess response and recovery needs have focused on collection of in situ data and direct reports from those situated on the ground during a flood event. This focus is often embedded within national- and local-level disaster management

acts and policies that were enshrined into law prior to recent advancements in Earth observation (van Niekerk, 2014). For response and recovery efforts, acts and policies have typically supported the development of cascading systems that rely on rapid collection of on the ground data and reports, participation from many local stakeholders, and strong coordination mechanisms to ensure that information can effectively flow upwards to higher levels of government. While in situ and locally reported data provide a strong foundation, there is potential to supplement the task of data collection immediately after a flood event through means of remote sensing and advanced modelling techniques.

Real-time data that monitors water flow, precipitation, and water level are critical for effective decision-making across the disaster risk management cycle (Sunkpho and Ootamakorn, 2011). However, conventional hydrological monitoring and network systems that rely on in situ data collection are often characterised by insufficient spatial and temporal coverage, limiting their application towards flood forecasting, mapping, and emergency response (Van Dijk and Renzullo, 2011). This challenge is exacerbated in many developing countries where the capital and maintenance costs of rain and stream gauging stations may limit the scale of data collection (Klemas, 2015). Certain types of satellite Earth observation data have emerged as potential cost-effective supplements or alternatives to in situ observations in data-sparse or entirely ungauged regions as the volume and resolution of data have improved (Khan et al., 2010; Rahman and Di, 2017). Satellite Earth observation data can be utilised not only for flood detection, but also to improve flood modelling by incorporating high-resolution digital elevation models and complementing hydrologic data assimilation (Revilla-Romero et al., 2015; Schumann, 2015; Tanaka et al., 2019). Significant research has emerged since the late 1990s to better understand

mechanisms in which remote sensing can be applied towards flood monitoring, flood risk management, and ultimately towards informing risk reduction measures (Schumann, 2015).

Both optical and microwave sensors have shown to be effective for measuring surface water (Huang et al., 2018). Optical sensors, such as those utilised by Landsat series of satellites and Sentinel-2, collect data across a variety of spectral bands ranging from the visible spectrum through shortwave infrared (SWIR). These types of sensors are passive; they capture the solar reflectance of the Earth's surface or atmosphere, resulting in the sensor being unable to penetrate cloud cover. Because major flood events are typically characterised by significant cloud coverage, optical sensors are often not effective for flood detection (Clement et al., 2018). Microwave sensors are active and can provide a viable workaround to challenges faced by cloud cover. Notably, the reflected energy from an object of interest can be measured when targeted by devices that emit radar signals.

Active sensors that collect Synthetic Aperture Radar (SAR) data have increasingly proven effective for monitoring changes in surface water (Musa et al., 2015). Notably, these technologies are effective in all-weather conditions, both day and night, and they can also detect standing water beneath vegetation canopies, an important attribute for flood inundation over densely vegetated land (White et al., 2015; Martinis and Rieke, 2015; Martinis et al., 2018). SAR-based sensors record the backscatter characteristics and roughness of different surface features. Because open water bodies typically exhibit lower backscatter coefficients, stagnant water tends to appear darker in the recorded image compared to non-water areas, which exhibit a higher backscatter (Brisco, 2015; Martinis and Rieke, 2015; Anusha and Bharathi, 2019; Cao et al., 2019). Previous studies rely on high-resolution SAR data collected

by TerraSAR-X (Wohlfart et al., 2018; Mason et al., 2009; Giustarini et al., 2012), RADARSAT-2 (Brisco et al., 2013; White et al., 2014), COSMO SKYMed (Pulvirenti et al., 2011; Pulvirenti et al., 2012) and ALOS PALSAR (Martinis and Rieke, 2015; Yulianto et al., 2015).

Because these satellites are commercial, the collected data is proprietary and its use for large scale analysis is limited. Sentinel-1 SAR data presents a free alternative and is rapidly becoming attractive for flood mapping across large geographic areas. Sentinel-1 provides routine, day-and-night, all-weather observation capability, making the data particularly suitable for emergency response support (Torres et al., 2012; Nasirzadehdizaji et al., 2019). Previous studies have shown the potential of Sentinel-1 data for flood detection (Martinis, 2017; Bioresita et al., 2018; Clement et al., 2018; Tsyganskaya et al., 2018; Uddin et al., 2019). However, the detection of inundation in complex environments, such as vegetated and urban areas, remain challenging due to the complexity of the radar signatures of such targets and various factors such as atmospheric conditions and high winds that induce surface roughening. Despite these constraints, there is an opportunity to better understand the extent to which SAR data, particularly freely available SAR data from Sentinel-1, can help improve decision-making around disaster risk management in data-sparse regions.

2.6. Conclusion

This chapter presented key definitions and concepts relevant to disaster risk management, outlined the evolution of disaster risk management as a global priority, and highlighted the critical importance of first understanding disaster risk before

making decisions across the disaster risk management cycle. Next, the chapter reviewed key technological and methodological advancements in areas of satellite Earth observation, localised data collection, and modelling techniques that have progressed our understanding of disaster risk. Lastly, the chapter focused on three key breakthroughs in disaster risk information for developing countries: (i) mapping population exposure; (ii) demonstrating more granular insights into population vulnerability; and, (iii) measuring the extent of flood hazard events.

Each of the technological and methodological areas discussed in this chapter, along with supporting literature, have advanced rapidly in recent years and will continue to do so in the years to come. Research in this space must remain cognisant of the cutting edge, while providing continuous examination of the opportunities and constraints of applying novel forms of data towards practical applications. The subsequent chapters of this thesis investigate the research questions posed, each of which pertains to specific aspects of a recent key breakthrough in disaster risk information discussed in this chapter. The body of research presented in this thesis builds upon existing literature to evaluate the validity and effectiveness of emerging techniques for high-resolution geospatial analysis applied to disaster risk management.

3. The utility of built environment geospatial data for high-resolution dasymetric global population modelling

Several global gridded population data sets are available at unprecedented high-resolution as demonstrated from recent efforts at 100-meter, 30-meter, and 10-meter resolution. These data sets are the result of the application of advanced methods to disaggregate census population counts from administrative units and facilitated by the proliferation of increasingly high-resolution spatial information pertaining to the built environment (e.g. built-up and building footprint data). Accordingly, these gridded population data are increasingly dependent on a single ancillary data set to inform the distribution of populations across space. Our study applies a flexible equation for high-resolution global dasymetric population modelling to test several combinations of binary masking variables (land areas, all building footprints, residential building footprints) and density variables (building footprint areas, building volumes) derived from characteristics of the built environment. The assessment is applied in New York City, where large spatial heterogeneities exist across confined geographic areas. Results confirm that the performance of methodologies generally improve as: (i) the binary masking variable becomes increasingly limiting; and, (ii) the density variable becomes more pronounced. However, application requires careful consideration due to their propensity to amplify both positive results and errors.

Keywords: population modelling, built environment, demography, remote sensing

3.1. Introduction

Robust population data represent an essential component of informed policy decision-making in areas including economic development, disaster risk management, transportation, and the environment. The collection of population data at a national level typically occurs through censuses and targeted household surveys. However, the frequency of a national census is typically only once per decade and household surveys often exist for a limited number of geographic areas and require a significant cost for data collection (Kilic et al., 2017). This highlights an important trade-off between accuracy and spatial resolution vis-à-vis currency and coverage that must be considered before these data can be applied at the local level.

High-resolution spatial data and statistical models that enable the mapping of population characteristics into smaller geographic areas are being developed in response to the demand for local-level statistics. Statistical approaches, such as dasymetric modelling, interpolation, cross-entropy, machine-learning and small area estimation, can help address the problem of providing feasible estimates. Several recent efforts deploying advanced methods have produced global and continental-extent gridded population data, which are becoming increasingly popular among applied scientists and practitioners (Leyk et al., 2019). Gridded population data have the desirable property of facilitating data integration with a consistent size (e.g. Patel et al., 2017). In addition, advancement in modelling techniques, data access and increased data quality have also led to an expansion of applications such as targeting vulnerable populations for welfare assistance (Watson et al., 2017), disaster risk management (Gunasekera et al., 2015, Ehrlich et al., 2018), urban development (Roberts et al., 2017), environment (Maroko et al., 2019) and climate change (Rigaud et al., 2018).

Despite efforts by national censuses to collect information for every individual and household, ethical constraints pertaining to the potential risk of identification often require data to be released publicly at an aggregate geographic level to preserve a certain degree of anonymity (see Langford and Harvey, 2001; Arzberger et al., 2004) and/or a sample of the total population (e.g. Integrated Public Use Microdata Series – IPUMS). The applicability of spatial analysis and modelling techniques therefore remains limited to the geographic level of release unless a form of spatial disaggregation is employed to estimate the distribution of populations and their associated attributes within aggregated areas. While it is possible to estimate populations counts by employing bottom-up, survey-driven approaches (Weber et al., 2018) or methods using mobile phone data (Deville et al., 2014), most high-resolution techniques utilise population censuses where available as both a baseline for spatial disaggregation, and as a measure to calculate the accuracy of the technique employed (Wardrop et al., 2018). For consistency in global population datasets, census data are often projected from a given year to the reference year. One method to distribute populations is areal interpolation, which refers to the transformation of geographic information from one source area to another target area (e.g. Tobler et al., 1997; Liverman et al., 1998, Mennis, 2003; Wu et al., 2005).

Even basic methods of areal interpolation hold an advantage over traditional choropleth mapping where more spatially detailed information may explain part of the variation in the underlying statistical surface (Wright, 1936; Goodchild and Lam, 1980; Lam, 1983). Methods such as areal weighting or inverse distance weighted interpolation assume that the variable of interest is distributed homogeneously within source areas (Nagle et al., 2014). Thus, for population modelling it is assumed that the probability that a person is living in a given location is always greater than zero.

Wright (1936) observed these techniques to be unfit for mapping a population surface and better suited for topography due to human settlement not being a continuous phenomenon.

Dasymetric modelling is a subtype of areal interpolation that approaches the population surface as discrete with sharp boundaries identified using ancillary variables (Mennis, 2003). Population allocation in dasymetric modelling is aided by two types of ancillary variables: limiting and related (Nagle et al., 2014). Limiting ancillary variables represent the constraints of the distribution whereas related ancillary variables represent the complex relationships between various ancillary variables where we know populations inhabit. For example, location of water bodies would be a limiting variable as one can assume those areas to be uninhabited.

Alternatively, proximity to roads is a related variable. One can conclude that people are more likely to live near a road rather than farther away. However, the relationship between where people live, and the location of the road, is undefined. The primary challenge historically for dasymetric modelling has been to define the relationship across a surface between the distribution of population and the set of ancillary variables from the multitude of possible specifications (Mennis, 2009). The most basic form of dasymetric modelling is the application of a “binary mask” comprised of one or several limiting ancillary variables (binary masking variables) where all known uninhabited areas are removed from consideration before the redistribution of populations is proportionally allocated across the remaining surface (Langford and Unwin, 1994). A step further in terms of complexity would be adding a density-informing ancillary variable to inform the spatial heterogeneities of the remaining population surface.

Uncertainty in output datasets is inherently difficult to quantify and compare due to a lack of validation data. However, one constant across all dasymetric modelling techniques is that the level of uncertainty increases as the source area becomes proportionally larger than the target area (Sadahiro, 2000). A common technique for reducing uncertainty, particularly important for higher-resolution population modelling, is the incorporation of many ancillary variables (Nagle et al., 2014).

To meet the growing demand for higher resolution population data, organisations and researchers have released gridded population layers at varying scales and levels of resolution. Using a raster data model allows for easy aggregation into areal units of any size, which facilitates integration with other raster data, and presents a better cartographic representation of density by standardising the size of each cell.

Table 4. *Prominent gridded population data sets that cover multiple countries by first year of release*

First year of release	Data set	Resol-ution	Method	Source	Reference Link
1995	GPW v1	10 km	Pycno-phylactic	Tobler et al., 1997	sedac.ciesin.columbia.edu /data/collection/gpw-v1
1998*	LandScan	1 km	Smart interpolation	Dobson et al., 2000	landscan.ornl.gov
2000	GPW v2	5 km	Areal weighting	Deichmann et al., 2001	sedac.ciesin.columbia.edu /data/collection/gpw-v2
2004	GRUMP v1	1 km	Dasymetric	Balk et al., 2005	sedac.ciesin.columbia.edu /data/collection/grump-v1
2013**	WorldPop	100 m	Statistical/ dasymetric	Stevens et al., 2015	www.worldpop.org
2015***	GPW v4	1 km	Areal weighting	Doxsey-Whitfield et al., 2015	sedac.ciesin.columbia.edu /data/collection/gpw-v4
2016	GHS-POP	250 m	Dasymetric refinement	Freire et al., 2016	ghsl.jrc.ec.europa.eu/ghs_ pop2019.php
2016	HRSL	30 m	Binary dasymetric	Tiecke et al., 2017	www.ciesin.columbia.edu /data/hrsl
2019	WSF	10 m	Binary dasymetric	Palacios-Lopez et al., 2019	N/A

*LandScan is open access for research purposes

**WorldPop united AfriPop, AsiaPop, and AmeriPop projects which had already released 100-metre resolution data sets prior to 2013 (Linard et al. 2012)

***Have not included GPWv3 as it did not improve the resolution of release or significantly change upon the methodology from GPWv2

Source: Balk et al. (2005) and Leyk et al. (2019)

One area of potential disruption to traditional approaches of disaggregating population counts is the large-scale detection of built-up land cover characteristics (Nieves et al., 2017), such as building footprints from either optical and/or synthetic aperture radar satellite imagery (Forget et al., 2018; Aamir et al. 2019) and building heights from synthetic aperture radar (Koppel et al., 2017) or UAV within urban areas (Rebelo et al., 2019). The Global Human Settlement Layer (GHSL) takes advantage of remotely sensed building footprint information identified through a Symbolic Machine Learning (SML) calculator and builds it up to a human settlement layer at 38-meter resolution (Pesaresi, 2018). The GHSL provides the base layer for GHS-POP, which produces population estimates at up to 250-meter resolution by combining the information with population estimates from GPWv4. The World Settlement Footprint (WSF) improves upon the GHSL by increasing the resolution to 10-meter using 2014-2015 Landsat 8 and Sentinel 1 optical and radar imagery (Strano et al. 2020) and shows promise for use in high-resolution population modeling (Reed et al., 2018; Palacios-Lopez et al., 2019). Similarly, the Global Urban Footprint (GUF) provides a layer of human settlement at 12-meter resolution (75-meter resolution for public usage) using images from 2010 to 2013 from radar satellites TerraSAR X and TanDEM X (Esch et al., 2011; Esch et al., 2017; Esch et al., 2018). Tiecke et al. (2017) highlight techniques for direct large-scale identification of building footprints using convolutional neural networks. Expanding from the building footprint layer into a population dataset, Tiecke et al. (2017) develop the High-

Resolution Settlement Layer (HRSL) at 30-meter resolution for 18 countries. The population datasets utilise a binary mask, then distribute population from the highest resolution available census administrative data evenly across discrete building footprints, which are then aggregated into gridded cells. Reed et al., (2018) analyse the relative effectiveness of several built-up land cover layers as inputs for three types of population models: (i) binary dasymetric; (ii) random forest; and, (iii) hybrid (first a binary mask, then redistribution using random forest). The random forest and hybrid models performed comparably, while both outperformed the binary dasymetric model. Stevens et al. (2020) also used several built-up land cover layers to test a binary dasymetric model against a random forest-based approach. The random forest-based model outperformed the binary dasymetric model in seven out of the 11 countries in the study.

The HRSL represents a shift away from a longstanding trend of increasingly complex techniques for modelling population surfaces, by favouring a simpler technique that generates a precise binary mask using only building footprints as a limiting ancillary variable. The HRSL then applies a density variable based off building footprint counts to inform the distribution of populations across the remaining population surface. For the HRSL, the same dataset provides the information to generate a binary mask and to inform (population) density. Tapp (2010) and Zandbergen (2011) provide a strong rationale to pursue this further by highlighting the robust performance of residential address points for predicting population counts. While building footprints may not prove as robust, they have the appeal of being derived from earth observation as a consistent measure across boundaries at scale as well as its parsimonious simplicity, which does not confound

subsequent modelling and does not depend on several ancillary variables (see Dmowska and Stepinski, 2017; Zoraghein and Leyk, 2019).

With recent advancements novel attribute classes for building footprint datasets are available to further refine density estimates. For example, lidar remote sensing enables collection of building height and volume data which were employed by Sridharan and Qiu (2013) as ancillary variables for population modelling. Furthermore, machine learning methods associated with lidar data have resulted in successful classification of building type, including single-family houses, multiple-family houses, and non-residential buildings, to over 70 percent accuracy (Lu et al., 2014). Kang et al. (2018) present a method to classify building types by combining satellite-derived building footprints with façade structures identified through the training of convolutional neural networks on street view images. On-the-ground surveying of building attributes such as typology and building height are less technologically advanced but can be quite effective and are often already existing and available for free through OpenStreetMap (OSM). OSM data and other on-the-ground surveys also show potential for use as training sets and validation for machine learning algorithms aimed at the widespread detection of building typology through earth observation. Tiecke et al. (2017) use GPS locations from a nationally representative sample of households in Malawi to validate building identification, which has the advantage of a sample that is independent of remote sensing. Although inputs derived from many localised techniques are not currently applied towards global population models, some may reach scale (e.g. global database of street view imagery) or achieved through earth observation data (e.g. building typologies in OSM) in the coming years.

In this context, we see two distinct objectives moving forward for advancing global dasymetric population models: 1) increasing the precision and accuracy of a limiting variable to constrain the spatial extent of settled populations; and, 2) improving modelling techniques of ancillary variables to account for variations in density. Given the increased emphasis that certain models place on a single ancillary variable to both limit the spatial extent of populations and inform density (e.g. HRSL with discrete building footprints), it is important to better understand how the relationship between the two may impact model performance. The main contribution of our paper is to test the accuracy of several combinations of binary masking and density variables potentially applicable towards the next general of global dasymetric population models in an environment with extreme variations in local population density.

3.2. Methodology

In this study, we apply a flexible equation for global dasymetric population modelling that separates a binary masking variable from a density variable for the purposes of assessing the accuracy of variables derived from characteristics of the built environment. Building on seminal methods of population modelling (Langford and Unwin, 1994; Mennis and Hultgren, 2006; Nagle et al., 2014), this equation specifically [frames/isolates/encapsulates] the two objectives that we have highlighted for advancement in global dasymetric population modelling: increasing the precision and accuracy of a strong limiting variable to constrain the spatial extent of settled populations; and, improving modelling techniques of ancillary variables to account for variations in density.

Each target cell c_i $(i = 1, \dots, m)$ within the broader source area is classified

with a binary masking variable b_i based on an assigned real or integer scalar attribute $a_{i,j}$ $(j = 1, \dots, n)$ where:

$$b_i = \begin{Bmatrix} 1: \sum_{j=1}^{n_i} a_{i,j} > 0 \\ 0: \sum_{j=1}^{n_i} a_{i,j} = 0 \end{Bmatrix}$$

And p represents the sum of all cells c_i where $b_i = 1$:

$$\sum_{i=1}^{m} b_i = p$$

So that expected population of the target cell $\hat{y}_i$ equals the sum of the population of all target cells y_s multiplied by the binary variable b_i (0 or 1) and divided by p, maintaining the pycnophylactic property:

$$\hat{y}_i = y_s \frac{b_i}{p}$$

We propose that the most scalable designation of b_i considering the present state of earth observation and image classification technologies utilises building footprint areas. However, b_i is envisaged to have flexibility engrained and can be defined as any variable that restricts the population to specific volumes, areas or points within the target cell c_i. This engrained flexibility enables for the incorporation of advancements which may further enhance the accuracy and precision of the binary mask such as the widespread detection of building typologies, which could be used to eliminate non-residential structures from consideration. Other examples of binary masking variables could include tax parcel records (Jia and Gaughan, 2016;

Zoraghein and Leyk, 2018) or address points (Zandbergen 2011).

The second piece of the equation is the density variable d_i, which acts to further refine the value of $\hat{y}_i$ to account for the fact that population is not inherently distributed evenly between each cell where $b_i = 1$.

Each target cell c_i is classified with a density variable d_i where:

$$d_i = \begin{Bmatrix} x: b_i \equiv 1 \\ 0: b_i \equiv 0 \end{Bmatrix}$$

And d_i is normalised so that:

$$\sum_{i=1}^{m} d_i = \sum_{i=1}^{m} b_i$$

So that:

$$\hat{y}_i = y_s d_i \frac{b_i}{p}$$

There are several methods to calculate the density variable, ranging from those based solely on attributes relating directly to b_i (e.g. sum of building footprint areas within c_i) to generating a basic function which relates distance from urban clusters (Tiecke et al., 2017) to more complicated methods that integrate many separate ancillary variables, such as the hybrid random forest approach used by Reed et al. (2018). through techniques including intelligent dasymetric mapping (IDM) (Mennis and Hultgren, 2006). The decision on which technique to use remains highly dependent on the type and accuracy of ancillary data available and the spatial resolution required for the fitness-for-use of the application (Leyk et al., 2019). However, one can expect that as the spatial variation of populations becomes

increasingly precise and accurate, the importance of attributes relating directly to b_i will increase correspondingly.

Fig. 3 shows the areal interpolation of a source area where $y_s = 80$ using three different techniques, which increasingly provide more detailed information regarding the population surface. The simplest is areal weighting, which when employed on a grid provides an even distribution of y_s across cells. The binary mask introduces ancillary data in the form of building footprints, which provides an even distribution of y_s across cells where $a_i > 0$. Lastly, the combination of the binary mask and the density variable allows us to show variance between populated cells.

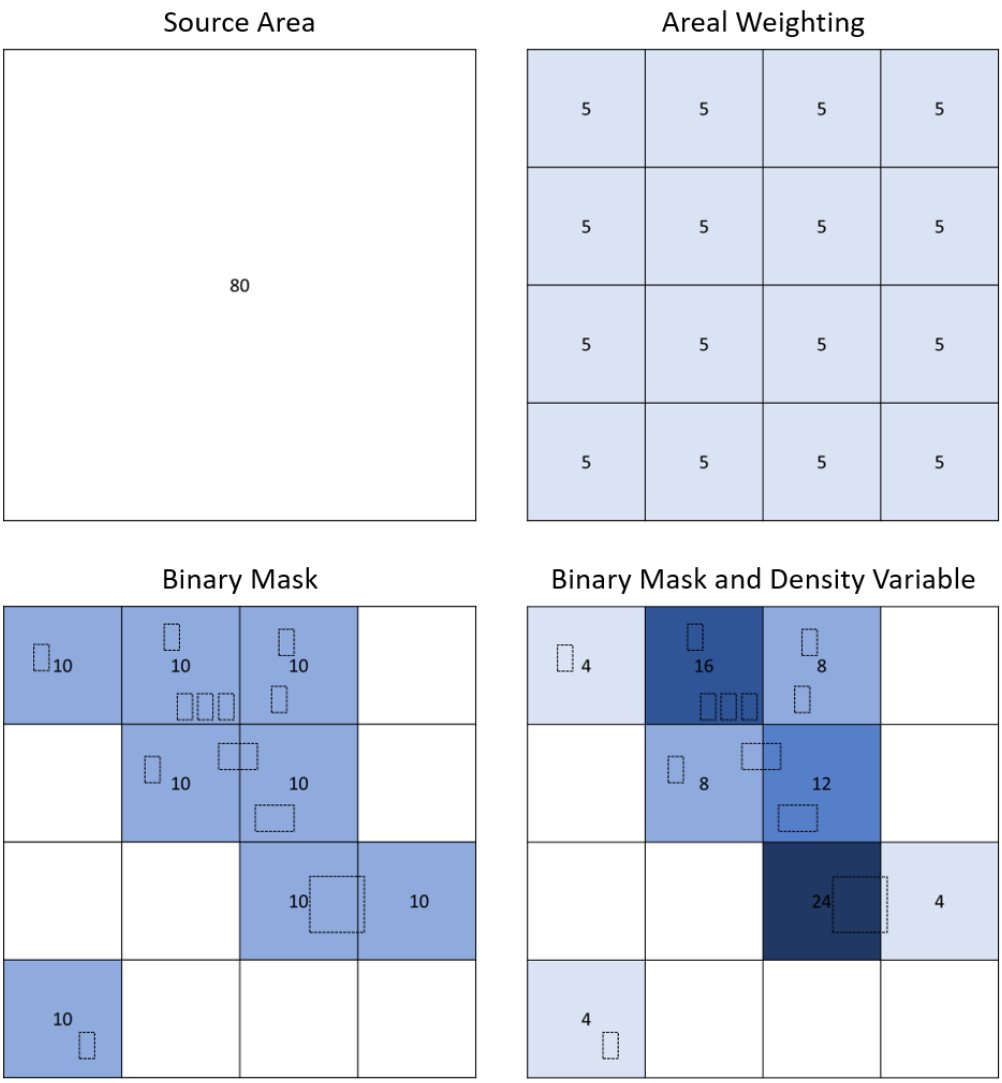


Fig. 3 Areal interpolation techniques for a source area with a population y_s of 80

3.3. Cast Study: New York City

Our study area focuses on New York City. Although New York City has characteristics that are unique with regards to geography, economic development, and zoning, it provides an opportunity for insight from a data-rich urban environment with extreme variations in local population density and diverse types of input data in which to construct multiple methods used in global population modelling. The New York City context presents an advantage over a synthetic analysis by forcing one to confront certain real-world data limitations (e.g. years may not align on all data sets)

and semantic errors (e.g. how to classify mixed use zoning) that are important to understand if one is to apply these variables in a less data robust setting. In this study area we compare the efficacy of multiple methodologies which apply a binary masking variable paired with a density variable derived from the same binary masking variable. We test the following methodologies:

1. Land areas binary mask
2. Building footprint binary mask
3. Building footprint binary mask and area density variable
4. Building footprint binary mask and volume density variable
5. Residential building footprint binary mask
6. Residential building footprint binary mask and area density variable
7. Residential building footprint binary mask and volume density variable

As a base to compare the effectiveness of subsequent methodologies, we start by applying a basic binary mask, which limits populations to only land areas – an important limiting variable considering the vast shorelines of New York City and one used in the GPWv4. We then focus on building footprint data to construct binary masking variables. We consider a set/range of increasingly limiting binary masks from all building footprints to only residential building footprints. We then consider a set/range of increasingly granular density variables from the same building footprint data, including building footprint area and building volume. Comparing increasingly limiting binary masking variables and increasingly granular density variables relating to the same underlying dataset allows us to compare projected future advancements in

built environment data derived from globally available earth observation. The results aim to shed light on some of the trade-offs between precision and accuracy with methods that use the same underlying dataset to derive both a binary masking and density variable. It also allows us to test the semantic accuracy by labelling buildings residential or non-residential and its subsequent impact on our results.

3.3.1 Data

In total, our study area is 1,213 square kilometres and contains 8,175,133 residents. We define our study area as all areas covered by the 2010 census block polygons in Bronx, Kings (Brooklyn), New York (Manhattan), Queens, and Richmond (Staten Island) counties. The study area contains 2,167 census tracts with an average area of 0.56 square kilometres and an average population of 3,773, and 39,148 census blocks with an average area of 3 hectares (0.03) square kilometres and an average population of 209. Census units in New York City can differ in size dramatically due to large areas of water and large variations in density (see Table 5).

Table 5. *New York City population, land area, and population density by borough*

	Population	Total Area (sq. km)	Land Area (sq. km)	Land Area Density (pp./sq. km)
New York City	***8,175,133***	***1,213.37***	***783.83***	***10,947***
Bronx	1,385,108	148.87	109.04	13,231
Brooklyn	2,504,700	251.02	183.42	14,649
Manhattan	1,585,873	87.00	59.13	27,826
Queens	2,230,722	461.56	281.09	8,354
Staten Island	468,730	264.94	151.18	3,132

A summary of the data layers used in the analysis can be found in Table 6.

Table 6. *Data Layers for New York City Analysis*

Type	***Variable Name***	***Description***	***Date***	***Source***
Vector	Building Footprints	Polygon features of each building footprint in Manhattan 95 percent of data accurate to +/- 2 feet – includes roof height attributes derived from TIN used for volumetric calculations	Source imagery from June 2014	New York City Department of Information Technology and Communications
Vector	Land Use	Polygon features of each tax lot – includes attribute for land use used for delineating residential buildings – mixed-use residential classified as residential for our study	May 2018	MapPLUTO 18v1, New York City Department of City Planning, Information Technology Division
Vector	Census Blocks for New York City Counties (both clipped to shoreline and not clipped to shoreline), FIPS Codes 005, 047, 061, 081, 085	Polygon features of each census block – includes attribute for population counts	2010	United States Census Bureau
Vector	Census Tracts for New York City Counties (both clipped to shoreline and not clipped to shoreline), FIPS Codes 005, 047, 061, 081, 085	Polygon features of each census tract – includes attribute for population counts	2010	United States Census Bureau

One should note that there is a gap of four years from 2010 to 2014 between the census data used as a baseline for disaggregating populations from administrative units and the building footprint data used to define variables in the dasymetric model. There is a further four years from 2014 to 2018 between the building footprint data and the land use data used to determine the residential areas for selected analysis. Although this represents a potential limitation of the study, because the purpose of the study is to compare methodologies relative to one another it was not deemed necessary to project populations into the future. Furthermore, while localised zoning changes certainly affect the accuracy of our determination of residential areas, there

were no major citywide changes to residential zoning policies between 2010 and 2018.

3.3.2 Analysis and results

Our study analysed the accuracy of each of the seven methodologies at two geographic levels: once defining the source area a_s as counties (boroughs) and once defining the source area a_s as census tracts. To assess the accuracy of the outputs, population totals of each 30-metre grid cell were re-aggregated into census blocks, the finest-level administrative unit available and our predicted population counts were compared to observed populations counts from the 2010 United States Census (see Fig. 4). The average census tract is roughly 20 times the size of the average census block, while the average county (borough) is over 8,000 times the size of the average census block.



Fig. 4 Census counties (boroughs), tracts and blocks at the intersect between Manhattan, Queens and Brooklyn boroughs

Within the study area, we generated 1.35 million grid cells at 30-metre resolution, of which 569,000 contained at least a partial building footprint, and 460,000 contained at least a partial residential building footprint. Within the study area, we generated 1.35 million grid cells at 30-meter resolution in order to align with HRSL, of which 569,000 contained at least a partial building footprint, and 460,000 contained at least a partial residential building footprint. Fig. 5 shows a 30-metre resolution gridded overlay of land use (labelled by land use code) and building footprints (labelled by building height in meters). Table 7 shows the land use codes, their description, and

whether they were classified as residential areas for the purposes of our analysis. Note that we have classified mixed-use as residential to ensure that all potentially inhabited areas are included in the residential building footprint binary mask.



Fig. 5 Land use (labelled by land use code) and building footprints (labelled by building height in meters) layers with 30-meter resolution gridded overlay near Columbus Circle, Manhattan

Table 7. *New York City land use codes and descriptions*

Value	Description	Residential
1	One- & Two-Family Buildings	Y
2	Multi-Family Walk-Up Buildings	Y
3	Multi-Family Elevator Buildings	Y
4	Mixed Residential & Commercial Buildings	Y
5	Commercial & Office Buildings	N
6	Industrial & Manufacturing	N
7	Transportation & Utility	N
8	Public Facilities & Institutions	N
9	Open Space & Outdoor Recreation	N
10	Parking Facilities	N
11	Vacant Land	N

Building footprint area and volume, as well as residential building footprint area and volume were calculated for each cell and applied to their relevant methodology. All work was conducted using ArcMap 10.6. Fig. 6 presents a map for the entire city utilising a residential building footprint binary mask and volume density variable downscaled from census tracts, while Fig. 7 demonstrates the increasingly enhanced spatial precision of each of the seven methodologies.

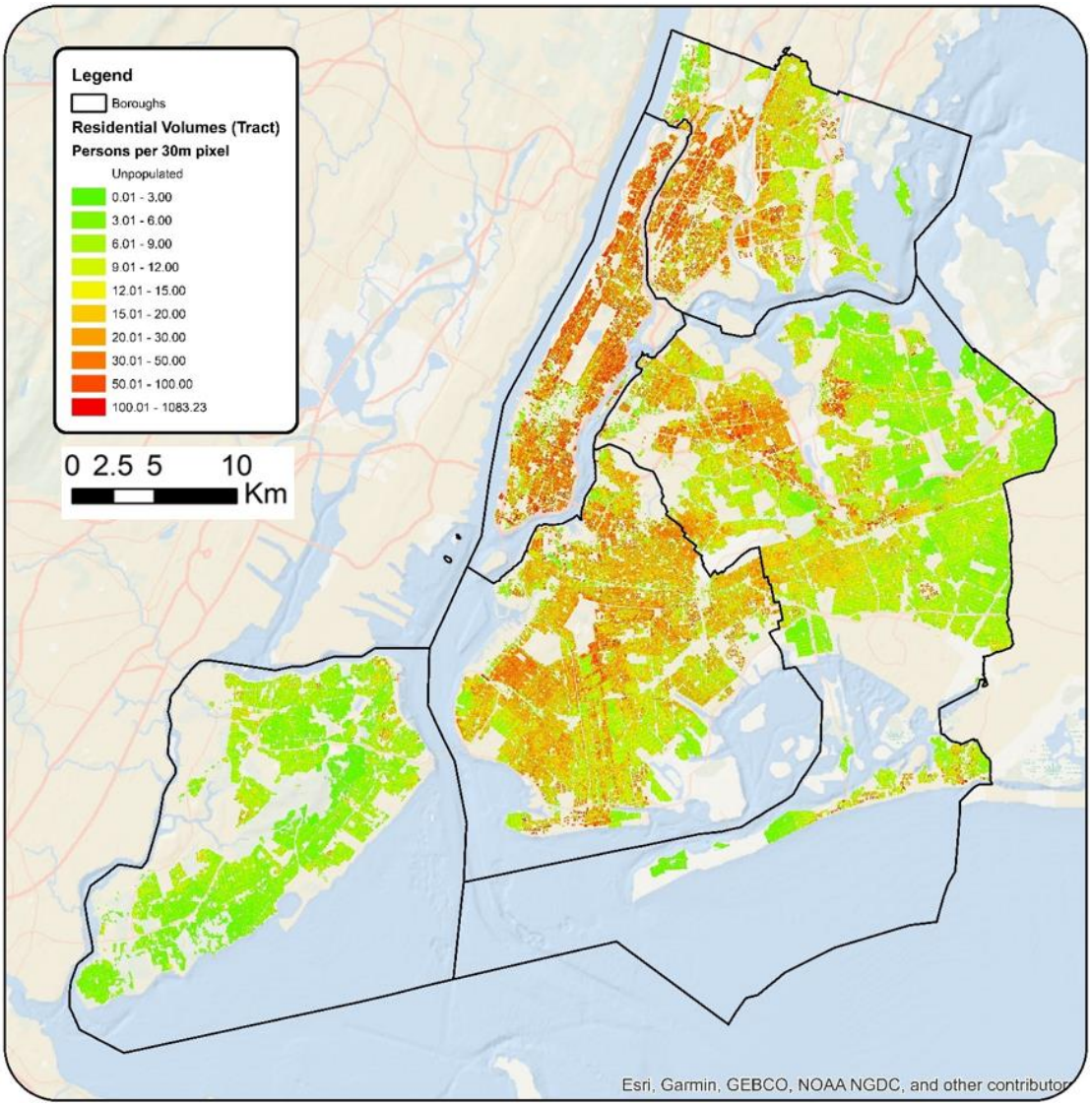


Fig. 6 New York City population density utilising a residential building footprint binary mask and volume density variable downscaled from census tracts

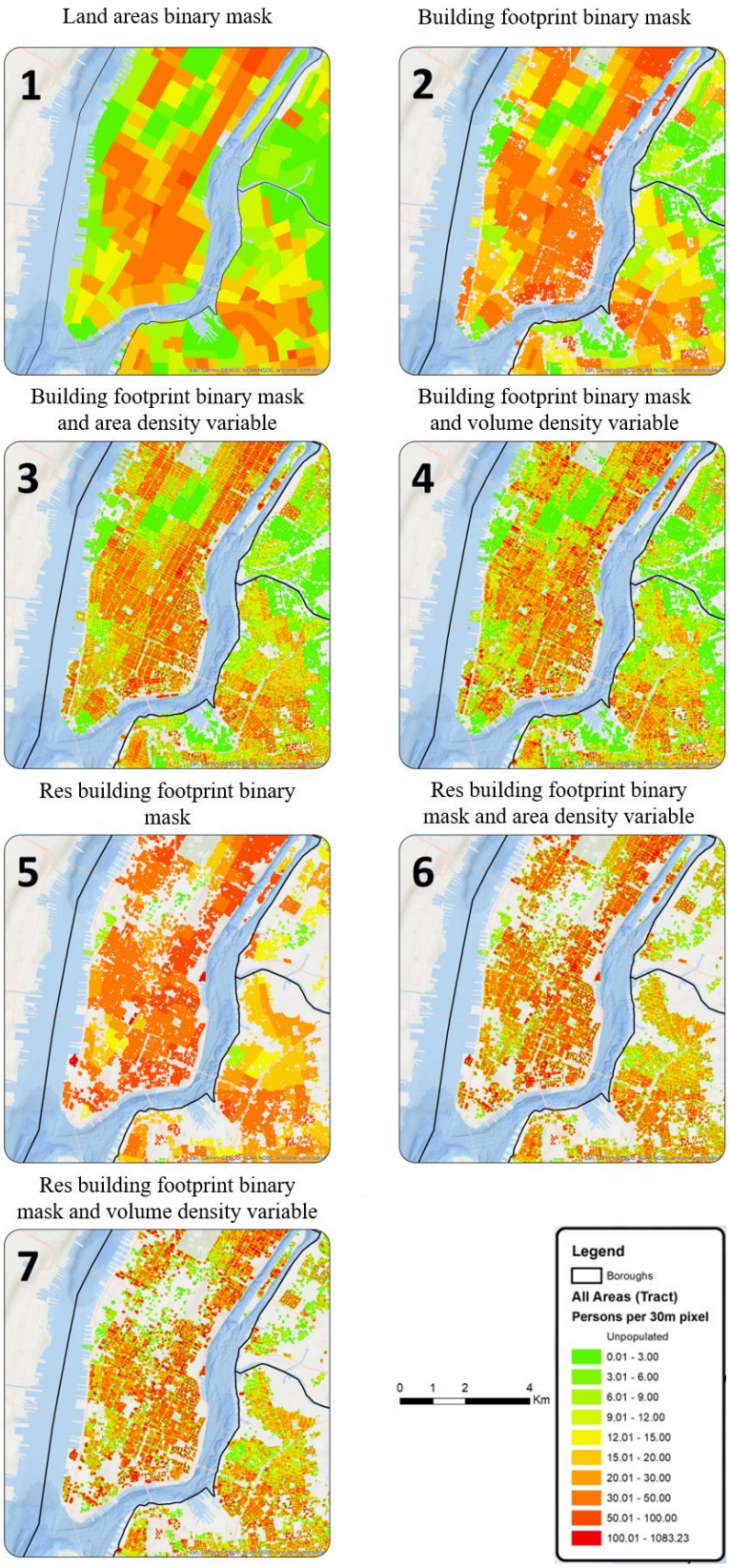


Fig. 7 Mapped results of the seven methodologies (20x downscaling) at the intersect between Manhattan, Queens and Brooklyn boroughs

The results of the accuracy assessments across the five boroughs for each of the methodologies are presented in Tables 8 and 9. We calculate the root mean square error (RMSE) and mean absolute error (MAE) for each methodology across all 39,148 census blocks. We also present the results as a percentage of the average population of census blocks within each borough, which are referred to as %RMSE and %MAE. The RMSE squares the errors before averaging them, meaning that it will increase with the variance of the frequency distribution of error magnitudes. In relation to MAE, RMSE gives a relatively higher weight to large errors as a result.

Table 8. *Borough-level disaggregation accuracy assessment (8,000x downscaling)*

Root Mean Square Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	843.44	869.19	882.00	811.46	302.36	792.84
BF Binary Mask	381.30	259.10	450.82	222.18	102.80	278.88
BF Areas	417.14	285.79	478.83	262.80	115.63	309.66
BF Volumes	460.17	362.90	704.92	406.77	146.71	422.27
Res BF Binary Mask	307.74	198.17	297.49	178.58	91.32	212.63
Res BF Areas	277.39	193.38	313.75	153.41	80.17	199.58
Res BF Volumes	255.65	214.30	542.14	156.89	69.48	245.47

% Root Mean Square Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	3.35	3.39	2.20	5.40	3.28	3.80
BF Binary Mask	1.51	1.01	1.12	1.48	1.11	1.34
BF Areas	1.83	1.11	1.19	1.75	1.25	1.48
BF Volumes	1.66	1.41	1.76	2.71	1.59	2.02
Res BF Binary Mask	1.22	0.77	0.74	1.19	0.99	1.02
Res BF Areas	1.10	0.75	0.78	1.02	0.87	0.96
Res BF Volumes	1.01	0.84	1.35	1.04	0.75	1.18

Mean Absolute Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	274.88	206.03	364.21	141.14	92.96	192.36
BF Binary Mask	205.29	148.96	270.49	102.99	46.14	138.35
BF Areas	191.65	146.32	290.08	93.57	43.43	133.83
BF Volumes	166.94	157.03	393.77	100.36	46.99	146.53
Res BF Binary Mask	172.45	114.59	181.69	91.80	39.64	111.11
Res BF Areas	144.31	97.62	178.77	72.51	32.87	94.43
Res BF Volumes	87.57	86.15	228.13	54.42	28.54	81.16

% Mean Absolute Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	1.09	0.80	0.91	0.94	1.01	0.92
BF Binary Mask	0.81	0.58	0.67	0.69	0.50	0.66
BF Areas	0.76	0.57	0.72	0.62	0.47	0.64
BF Volumes	0.66	0.61	0.98	0.67	0.51	0.70
Res BF Binary Mask	0.68	0.45	0.45	0.61	0.43	0.53
Res BF Areas	0.57	0.38	0.45	0.48	0.36	0.45
Res BF Volumes	0.35	0.34	0.57	0.36	0.31	0.39

Table 9. *Tract-level disaggregation accuracy assessment (20x downscaling)*

Root Mean Square Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	291.17	183.62	366.01	153.99	143.93	213.35
BF Binary Mask	273.23	259.10	450.81	222.18	102.80	181.15
BF Areas	297.23	285.79	478.83	262.80	115.63	196.84
BF Volumes	265.68	362.90	704.92	406.77	146.71	191.93
Res BF Binary Mask	228.18	134.66	244.63	104.86	68.36	150.51
Res BF Areas	220.73	132.23	257.81	95.61	63.63	148.00
Res BF Volumes	195.18	107.48	283.97	82.91	50.13	138.93

% Root Mean Square Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	1.16	0.72	0.91	1.03	1.56	1.02
BF Binary Mask	1.08	0.63	0.75	0.84	0.85	0.87
BF Areas	1.18	0.68	0.83	0.88	0.95	0.94
BF Volumes	1.05	0.63	0.87	0.90	1.07	0.92
Res BF Binary Mask	0.91	0.52	0.61	0.70	0.74	0.72
Res BF Areas	0.88	0.52	0.64	0.64	0.69	0.71
Res BF Volumes	0.77	0.42	0.71	0.55	0.54	0.67

Mean Absolute Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	152.47	112.19	225.91	72.06	58.59	107.13
BF Binary Mask	132.55	148.96	270.49	102.99	46.14	88.94
BF Areas	133.52	146.32	290.08	93.57	43.43	89.03
BF Volumes	112.30	157.03	393.77	100.36	46.99	84.27
Res BF Binary Mask	99.26	76.21	136.77	47.99	29.76	68.81
Res BF Areas	85.12	53.04	132.82	39.06	24.99	59.60
Res BF Volumes	63.23	64.86	133.77	31.96	21.35	50.51

% Mean Absolute Error

Methodology	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Land Areas Binary Mask	0.61	0.44	0.56	0.48	0.63	0.51
BF Binary Mask	0.53	0.58	0.67	0.69	0.50	0.43
BF Areas	0.53	0.57	0.72	0.62	0.47	0.43
BF Volumes	0.45	0.61	0.98	0.67	0.51	0.40
Res BF Binary Mask	0.39	0.30	0.34	0.32	0.32	0.33
Res BF Areas	0.34	0.21	0.33	0.26	0.27	0.29
Res BF Volumes	0.25	0.25	0.33	0.21	0.23	0.24

3.4. Discussion

In the following sub-sections, we present the methodology performance at the borough level, the relation of methodology performance to census block density and the relation of methodology performance to the coefficient of variation of census tracts.

3.4.1 Methodology performance at the borough level

Results by New York City borough present a broad picture of overall performance for each of the seven methodologies tested across real geographies with large localised variations in population density. However, a few key trends are noticeable:

1. *The relative effectiveness of the land areas binary mask is heavily linked to the degree of downscaling.* As demonstrated in Fig. 8, all tested methodologies were substantially more accurate at 20x downscaling (tracts to blocks) than at 8,000x downscaling (counties to blocks). However, the effects were significantly more pronounced with the land areas binary mask, particularly in terms of RMSE. At 20x downscaling, the RMSE and MAE of the land areas binary mask are both 1.2x the building footprint binary mask; at 8,000x downscaling, the RMSE of the land areas binary mask increases to 2.8x that of the building footprint binary mask while the MAE increases to 1.4x that of the building footprint binary mask. This is largely intuitive, but it highlights the utility of applying a strong binary mask where the degree of downscaling required is significant.

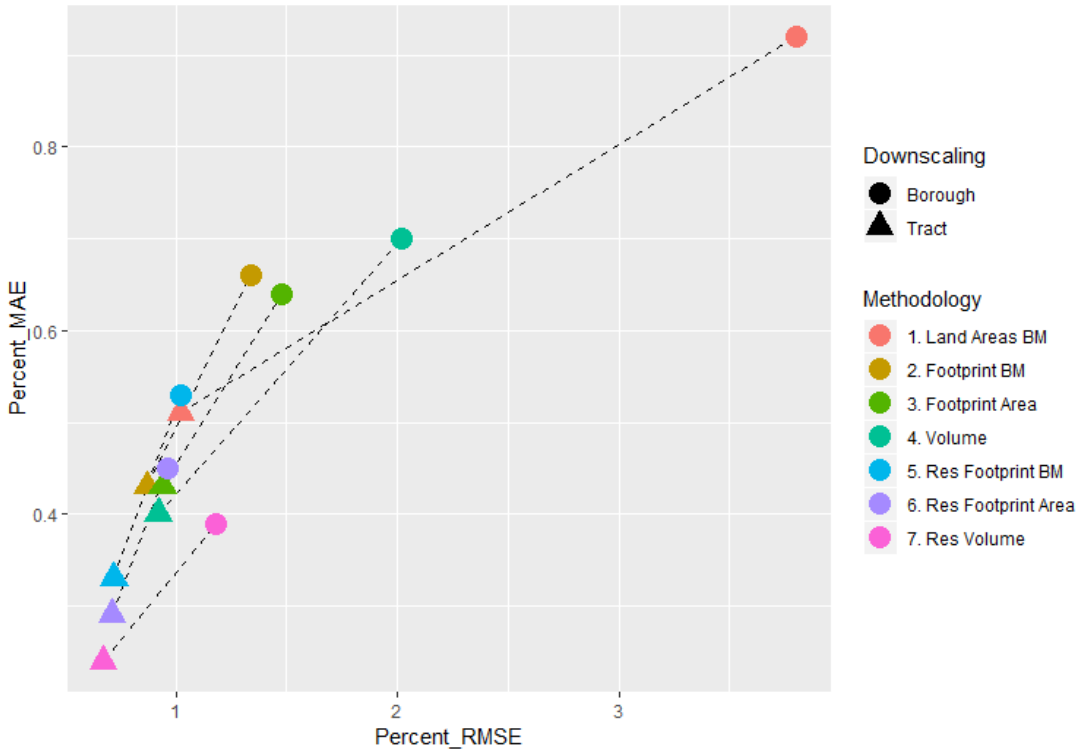


Fig. 8 %RMSE and %MAE for each methodology at 20x and 8,000x downscaling

2. *Improvements to the binary mask consistently improve accuracy.*

   Methodologies utilising residential building footprints rather than all building footprints performed universally better in terms of RMSE and MAE. For both 8,000x downscaling and 20x downscaling the RMSE for the all building footprint binary mask was 1.2x that of the residential binary mask, while the MAE was 1.3x. Fig. 9 provides borough-level comparisons.

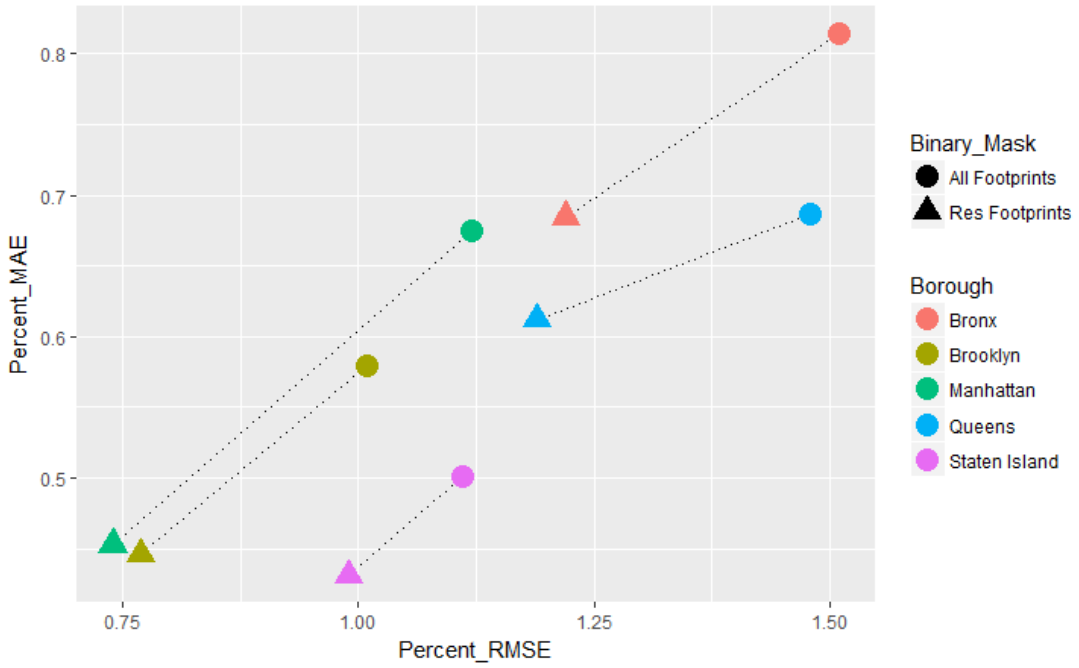


Fig. 9 %RMSE and %MAE for each borough with all building footprint binary mask and residential building footprint binary mask

3. *Mixed results overall for density variables.* Including building footprint areas and building volumes as density variables provided mixed results for the RMSE of the output layers at both 8,000x and 20x downscaling, while they provided marginally improved results for MAE. This makes sense if you consider that density variables inherently increase the variance of populations between cells in comparison to a binary mask, leading to the potential amplification of errors where they are present. In our case, we classified mixed-use residential as residential – an example of a potential semantic error, which could have large ramifications in certain target areas where mixed-use residential is more prevalent. In dense parts of Manhattan, for example, a semantic error in mislabelling of a large office building could have large ramifications. Despite this trade-off, the most accurate layer overall for both

RMSE and MAE was the residential building footprint binary mask with the building volume density variable.

3.4.2 Methodology performance in relation to census block density

While borough-level analysis provides the best indication of how each of the models perform at scale across a diverse population surface, we also wanted to test how the methodologies performed specifically in relation to population density of census blocks and how significant localised differences in population density within census tracts affect the accuracy of our models.

We started by classifying census blocks by population density per square kilometre d_b into six categories as seen in Table 10. We use the established minimum criterion of 1,000 persons per square mile (386.12 per square kilometre) within an urban core to qualify as an urbanised area defined by the United States Census Bureau (Ratcliffe *et al.* 2016). Then, we define low-density blocks that are less than the minimum criterion, medium-density blocks are up to 10x denser, medium/high-density blocks are between 10x and 30x denser, high-density blocks are between 30x and 100x denser, while extremely high-density areas are over 100x denser. Because we are primarily concerned with the relative effectiveness of density variables, we only analyse three layers at 20x downscaling: 1) residential building footprint binary mask; 2) residential building footprint binary mask with building footprint areas; and, 3) residential building footprint binary mask with building volumes. The results are in Table 11.

Table. 10 *Classification of census blocks by population density (persons per square kilometre)*

No	Classification	Census Block Density	Obser-vations	Average Pop
1	Unpopulated	$d_b = 0$	9,018	0
2	Low-density	$d_b > 0 \leq 386.12$	476	12
3	Medium-density	$d_b > 386.12 \leq 3861.2$	3,345	50
4	Med/high-density	$d_b > 3861.2 \leq 11583.6$	11,367	117
5	High-density	$d_b > 11583.6 \leq 38612$	11,628	231
6	Extremely high-density	$d_b > 38612$	3,198	836

Table 11. *Classification of census blocks by population density*

Root Mean Square Error

Method	Un-populated	Low-density	Med-density	Med/high-density	High-density	Ext high-density
Res BF Binary Mask	65.15	184.04	102.89	83.89	123.42	373.05
Res BF Areas	69.95	195.08	86.84	74.27	122.50	363.84
Res BF Volumes	87.79	204.89	88.14	70.98	113.02	328.49

% Root Mean Square Error

Method	Un-populated	Low-density	Med-density	Med/high-density	High-density	Ext high-density
Res BF Binary Mask	NA	14.97	2.05	0.72	0.54	0.45
Res BF Areas	NA	15.87	1.73	0.63	0.53	0.44
Res BF Volumes	NA	16.67	1.75	0.61	0.49	0.39

Mean Absolute Error

Method	Un-populated	Low-density	Med-density	Med/high-density	High-density	Ext high-density
Res BF Binary Mask	17.41	66.63	56.48	43.90	67.60	235.81
Res BF Areas	9.73	47.54	38.59	35.88	60.33	218.72
Res BF Volumes	11.54	48.08	33.94	29.16	49.01	188.93

% Mean Absolute Error

Method	Un-populated	Low-density	Med-density	Med/high-density	High-density	Ext high-density
Res BF Binary Mask	NA	5.42	1.12	0.37	0.29	0.28
Res BF Areas	NA	3.87	0.77	0.31	0.26	0.26
Res BF Volumes	NA	3.91	0.68	0.25	0.21	0.23

Increasing the intricacy of the density variable returns poorer results for %RMSE in unpopulated and low-density blocks while we observe improved performance for each of the remaining categories. %MAE is consistently better for methodologies containing density variables. This again shows that increasingly

intricate density variables magnify the distribution of populations to be more accurate overall, but at the cost of accentuating errors where they are present, as evident in low-density blocks. These errors can be significantly reduced through incorporation of ancillary data into the density variable where data are available. The %RMSE and %MAE of all methodologies generally improved as the density category was increased.

3.4.3 Methodology performance in relation to the coefficient of variation of census tracts

Lastly, we tested the same three methodologies to observe how significant localised differences in population density within a sole source area affect the accuracy of our models. To do so, we calculate the coefficient of variation, which involves dividing the standard deviation of census block population density within each census tract by the mean population density of census blocks within the census tract. The larger the value on the x-axis, the larger the variance in population density between blocks with a single tract. We then compare the coefficient of variation to the %MAE for the population of each census tract using census blocks as observations. Fig. 10 shows that %MAE increases dramatically for all methodologies as the coefficient of variation increases from 0 to 1, after which it stabilises and subsequently increases. Footprint areas performed better than the binary mask, while footprint volumes outperformed both, highlighting the importance of a strong density variable for source areas with elevated levels of localised variation.

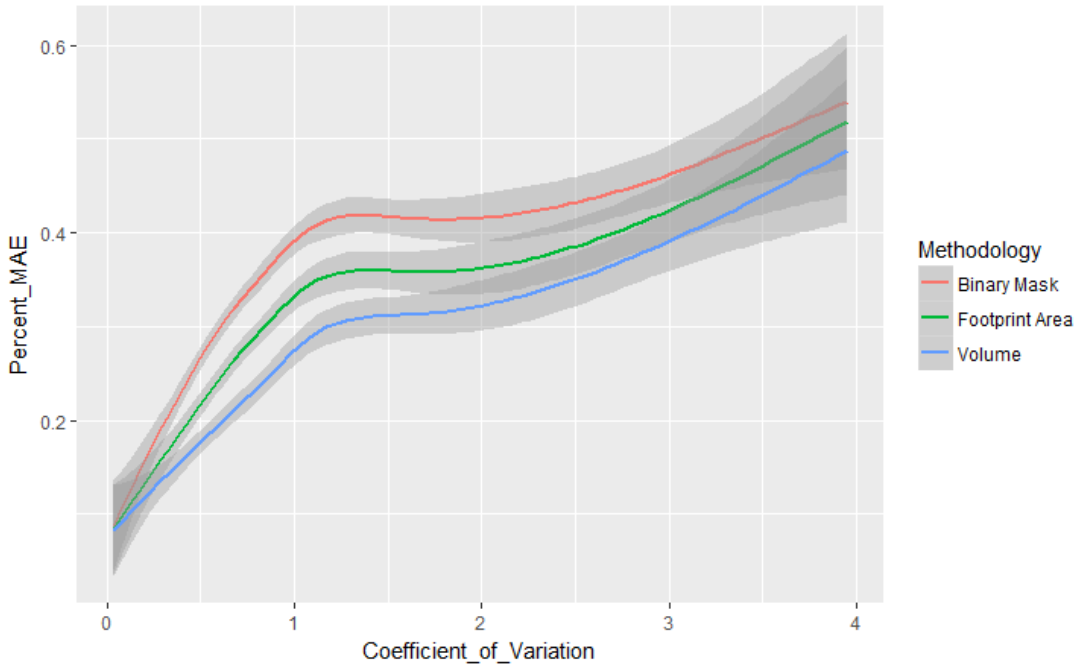


Fig. 10 Coefficient of variation and %MAE of census blocks within census tracts at 20x downscaling using GAM smoother

3.5. Conclusions

Our study analysed the effectiveness of applying increasingly limited binary masks and increasingly strong density variables directly related to the masking variable. Improving the precision of a building footprint binary masking variable can be linked to enhanced classification of building type, which has potential for refinement through several machine learning applications (e.g. OSM training sets, computer vision of street view imagery, rooftop size and characteristics from satellite imagery, business registers, census listing, etc.). Improving the accuracy of the binary masking variable can be linked to similar advancements in classification techniques focused on large-scale analysis of earth observation data. Our New York City case study presented a large, complete and verified building footprint and building heights dataset, as well as

zoning information in which to test our methods in a complex urban setting with high spatial heterogeneity. Although data derived solely from earth observation and machine learning applications would not currently compare in robustness, New York presented an opportunity to test our assumptions in a proxy environment considering the direction of future advancements.

Our results present evidence to pursue high-resolution dasymetric modelling techniques in environments without as strong of means for verification as New York City, where census blocks used for verification average only 3 hectares. The utility of a strong binary mask for downscaling of populations is particularly noticeable. The %RMSE and %MAE by borough at 8,000x downscaling utilising a residential building footprint was comparable to the %RMSE and %MAE by borough of 20x downscaling utilising a land areas binary mask. Our study also showed that increasing the precision of the building footprint binary mask to only include those which are residential or mixed-use improved our results by an average of 21 percent for RMSE and 29 percent for MAE. However, certain cases were much more pronounced, including that of Manhattan where a 115 percent improvement in MAE was observed at 20x downscaling with the residential building footprint binary mask. While it remains true that the level of uncertainty will always rise as the source regions becomes increasingly large in comparison to the target region, continued advancements in the precision and accuracy of a strong binary masking variable can substantially mitigate that uncertainty and extend the effectiveness of dasymetric population modelling techniques to support increasingly robust higher-resolution outputs.

Lastly, our study provided empirical insight into the performance of building footprint areas and building volumes across boroughs, block-level density classifications, and levels of density variation within tracts. Both building footprint areas and building volumes are density-informing ancillary variables directly relating to the masking variable. While our flexible equation is structured to accommodate all types of density variables, the purpose is to assess the increasing precision and accuracy of a strong limiting variable to constrain the spatial extent of settled populations; and, improving modeling techniques. Those directly relating to the masking variable demanded further analysis due to their increased usage for informing dasymetric population models including the HRSL, which utilises building footprint areas. Our conclusions confirm that density variables directly relating to the binary masking variable are effective however their application requires careful consideration due to its propensity to amplify both positive results and errors. We would thus recommend that to the greatest extent possible they be used in combination with additional density-informing ancillary variables not directly relating to the masking variable. This would serve to reduce the impact of large errors, which would in turn increase the effectiveness of the density-informing variable directly relating to the binary mask.

3.6. Data Statement

The census block error tables that support the findings of this study are available at https://figshare.com/s/1fd228faad6cef74d4ad. Land use data from MapPLUTO (Release 18v1) used to differentiate residential buildings can be found at https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page.

Building footprint data from NYC Open Data can be found at:

https://data.cityofnewyork.us/Housing-Development/Building-Footprints/nqwf-w8eh.

Census blocks, tracks, and counties data can be found at the TIGER/Line Shapefiles database: https://www.census.gov/cgi-bin/geo/shapefiles/index.php.

4. Spatially referenced synthetic households for high-resolution population modelling

This paper presents a methodology for generating spatially referenced synthetic households by combining population synthesis techniques for small area estimation with dasymetric population modelling. The methodology results in an output vector layer of households represented as points with unique structural and socioeconomic attributes which can be applied towards a better understanding of population dynamics, particularly those characteristics which are geographically influenced and may not conform to administrative boundaries linked to aggregated census statistics. To demonstrate the effectiveness of the methodology, the paper provides a case study in Dumuria Upazila, Bangladesh, where an independent household survey is used to validate the location of the synthetic households and associated socioeconomic attributes. The study found that 98 percent of all surveyed households were within 100 meters of at least one synthetic household and the median distance observed was only 18 meters. The study also found that the age distribution of synthetic populations was within five percent of the surveyed population for each of the five age bands considered.

Keywords: population modelling; population synthesis; geographic information systems; Bangladesh

4.1. Introduction

4.1.1. Anonymity and Geospatial Data

Population-based research can benefit significantly from detailed location information of the households or individuals studied (Olson et al., 2006). However, ensuring anonymity to a survey respondent is both essential for ensuring truthful responses and an ethical obligation for potentially sensitive data. Thus, most surveys and national censuses release their data for public usage in aggregate form, often paired with microdata that details out a fraction of anonymised responses for unique households (Arzberger et al., 2004). For example, the United States Census Bureau has disclosure avoidance requirements which restrict the release of microdata for public use to aggregate populations of at least 100,000 (Gomez, 2017). Generally, higher level aggregates diminish the effectiveness of small area estimation techniques, mask localised spatial dependencies, and may lead to ecological fallacies at lower levels of aggregation (Paiva et al., 2014). Anonymised household and individual level releases of microdata can help provide a baseline for understanding the distribution of attributes collected within a geographic area but may not provide much further detail into the spatial variance of those attributes. To address this concern, geographically focused releases of microdata with higher resolution geographic indicators have been released, often with more restricted attributes. Initiatives such as the Integrated Public Use Microdata Series (IPUMS) have worked to populate a database of individual-level microdata samples, including from international census records. Efforts such as the Demographic and Health Survey (DHS) program, the Living Standards Measurement Study (LSMS), and the Multiple Indicator Cluster Surveys (MICS) have also helped to provide baseline socioeconomic information in developing countries. Although an improvement, the effectiveness of household or individual

level spatial analysis and modelling remains limited without precise location data tied to features. Data have been used previously in this context but are typically dependent on either complete confidentiality or unique circumstances (Kramer et al., 2016).

The root of the challenge is an inherent trade-off between maximising the usefulness of the data, and safeguarding anonymity. Anonymisation aims to safeguard individuals, households, or larger groups of people from having potentially sensitive data attributed to them. This can include removal of information such as name, address, postal code, or any other attribute that could be used as a unique identifier. Assurances of anonymity are typically an important prerequisite for respondents to disclose potentially sensitive information. Removing location information is thus a key objective of anonymisation, making spatially focused research particularly difficult to pursue without some sort of workaround. Although government agencies such as the United States Census Bureau or the United Kingdom's Information Commissioner's Office have established guidelines for data anonymisation, a global consensus does not exist on the level of anonymisation required or the best technique to employ for each type of data (Brasher, 2018). This includes whether to release associated microdata for aggregate data sets, and in what type and quantity. Anonymisation of data is therefore largely a qualitative judgment, leaving spatial analysts and modelers in a position where each survey or census requires a tailored approach to understand what level of information can be extracted within a comfortable margin of error, and which methods for downscaling are most appropriate to employ given the context of the methodology and modes of data release.

4.1.2. Geospatially Referenced Surveys

Increasingly, geographic coordinates of features are being collected during surveys and censuses. Cheap, GPS-enabled smartphones and free, open source data collection tools such as Open Data Kit (www.opendatakit.org) have enabled even the simplest of surveys to include a high-resolution spatial element. The efficacy of these techniques was exemplified during the 2013-2016 West Africa Ebola epidemic, where the analysis of data collected though GPS-enabled smartphones helped response teams better predict, contain and treat outbreaks (Tom-Aba et al., 2015). Governments, particularly in developing countries where robust and up to date address points are not available, are increasingly turning to quickly mapping geographic coordinates as a means for identification. For example, following the 2015 Nepal Earthquake, a census of all houses within earthquake-affected districts was conducted which linked attributes of households, damaged structures, and photos to a precise geographic location using Android tablets (Lallement et al., 2017). Although this proved incredibly useful for the Government of Nepal to mobilise recovery and track reconstruction efforts at the household level, non-government organisations and the public were initially unable to access any data beyond simplified aggregates due to an inability to quickly determine an appropriate mode of anonymisation and release.

The question of how to effectively anonymise geospatially referenced data is becoming increasingly important as the collection of respondent-level spatial information becomes easier to implement and more standardised into mainstream methodology. In situations where the objective is to analyse point data for individual or household data collected from geospatially referenced surveys, techniques such as geographical masking (Zimmerman et al., 1999) and spatial blurring (Cassa et al.,

2008) can be employed to add layers of anonymity. Spatial blurring, for example, is a technique which moves point information a randomised distance from its true location. These techniques could theoretically be applied to a survey or census and would present the actual attributes collected by surveyors unless synthetic data are inserted as a replacement. Although access to geospatially referenced data would be ideal, unless involved in the collection process directly, the established norm is to only release data aggregated to a geographic resolution sufficiently diluted to ensure appropriate levels of anonymisation. This may change in the future, but for the time being it means that the more common workaround would require the downscaling, or disaggregation of data into smaller areas of higher geographic resolution.

4.1.3. Geospatial Workarounds to Data Aggregates

Small Area Estimation Techniques

At a basic level, techniques for downscaling socioeconomic data, including small area estimation (SAE), can be used to improve the resolution of aggregates available for comparison (Ghosh and Rao, 1994). Aggregated units are typically linked to census blocks or other administrative units but can also be linked to pre-drawn grid cells or other arbitrary boundaries. There is typically a trade-off between the number of attributes disclosed for each household and the geographic resolution in which the households can be identified. For example, an administrative level 2 microdata release may contain 20 attributes for each unique household, while an administrative level 4 microdata release may contain 10 attributes for each unique household. Usually the attributes that are being cut from the list are ones which are potentially sensitive if identified with the area of release.

Aggregate releases of data are often paired with microdata releases, which

provide individual or household attributes for a percentage of households (typically up to 10 percent) within the aggregated geographic unit. Paired with aggregates, microdata can be applied to SAE methodologies for downscaling. In situations where official microdata directly taken from the census records are not released or where attributes not collected in the census are required, surveys collecting some of the same or similar attributes can be used instead, or in connection with official microdata for the purposes of SAE. This is exemplified through the Elbers, Lanjouw and Lanjouw (ELL) method of SAE to derive estimates of poverty rates, which require supplemental attributes related to welfare not typically collected in a national census (Elbers et al., 2003). In Bangladesh, the World Bank (2010b) uses the ELL method of SAE developed to derive estimates of poverty rates at the district and *upazila* level. This is done by linking parameter estimates from a consumption model developed based on the 2010 Household Income and Expenditures Survey (HIES) with the 2011 Population Census.

Although many unique methodologies have been proposed for SAE (e.g. Bishop et al., 2007; Ghosh and Rao, 1994; Whitworth, 2013), they typically fall under one of two categories: statistical regression-based techniques or spatial microsimulation techniques. Statistical SAE techniques, such as the ELL method, match explanatory variables from the higher-level data set with those available from a separate data set in the smaller area of interest. This has proven a particularly robust technique for downscaling aggregates (i.e. administrative level 2 to administrative level 4) since results are inherently linked to confidence intervals around central point estimates. Statistical SAE techniques therefore concretely allow for uncertainty to be considered, and for downscaling to stop at a level where the analyst determines it is no longer of value. Although advancements have been made to calculate uncertainty

using certain spatial microsimulation techniques, it largely remains a methodological weakness of the technique when used for SAE (Whitworth et al., 2017).

Population Synthesis

The primary advantage of using a spatial microsimulation technique is that it results in the generation of a synthetic population for each target area. These synthetic populations can be utilised in distinct forms of modelling, including bottom-up approaches for understanding complex geographic systems. Recent years have seen substantial advancements in these areas, particularly with geographic automata systems (Torrens and Benenson, 2005), agent-based modelling (ABM), and spatial microsimulation (Lovelace and Dumont, 2016). ABM and spatial microsimulation techniques have demonstrated their use for policy decision making, where the synthetic datasets can be used as inputs to simulation models that evaluate the ex-ante distributional effect of proposed policies or development programmes. This has proven particularly effective in the transport sector, where agents are generated to be as reflective as possible of demographic structures and travel behaviour of real people (Voellmy et al., 2001). Other sectors where ABM has demonstrated potential include coupled socio-ecological systems (Filatova et al., 2013), and health (Tracy et al., 2018). Synthetic populations are particularly valuable for applying ABM techniques to analyse individual or household level interaction and decision-making. Improving the spatial resolution which populations can be disaggregated into agents therefore improves the level in which the agents are able to reflect the populations which they represent.

Creating a dataset of synthetic populations through spatial microsimulation techniques requires the application of algorithms to extract information from various

data sources to construct an entirely new dataset with the required attributes of households (Munnich and Schurle, 2003). Techniques for spatial microsimulation SAE can be clustered into three methodologies: 1) iterative proportional fitting (IPF); 2) combinatorial optimisation; and, 3) generalised regression reweighting (Whitworth et al., 2017). Although each method for spatial microsimulation SAE differs in substantial ways, they all "fit" the microdata to the multidimensional aggregates and generate representative synthetic populations. The raw synthetic population data can be useful on its own or be re-aggregated into the smaller area of interest. Standardised methodologies can be employed directly through statistical software, such as the R package simPop, an open source data synthesiser based on a modular object-oriented concept (Templ et al., 2017). Spatial microsimulation techniques for SAE first present a top-down approach, and then shift to a bottom-up approach. Regardless of which method for SAE is employed, the output remains aggregated in a spatial sense. While the release of anonymised microdata paired with aggregated data sets enables data scientists to investigate the relationships between collected attributes and increases the resolution in which SAE techniques can estimate spatial distribution, the spatial component often remains lost at a higher resolution.

Dasymetric Population Modelling

While SAE focuses on downscaling individual or household attributes, spatial population modelling focuses on the spatial distribution of individuals or households over a population surface. The primary difference between the two is that spatial population modelling focuses solely on the geographic distribution of the population over a discrete surface, while SAE techniques focus on the downscaling of population characteristics (e.g. socioeconomic attributes). As a result, both techniques rely

heavily on the disaggregation of census data but approach it from a distinct perspective. Dasymetric modelling, a branch of areal interpolation where boundaries between ancillary variables are discrete, is often employed for spatial disaggregation of population data. In many cases, there is little agreement on which subset technique to employ due to an inherent subjective element to determining the quality and importance of available spatial datasets, as well as the output resolution required. However, even the most basic methods of areal interpolation, such as areal weighting, hold an advantage over choropleth mapping for gaining insights into the spatial variation of populations (Wright, 1936; Goodchild and Lam, 1980; Lam, 1983).

Access to increasingly high-resolution free and open spatial data sets derived from earth observation (e.g. Sentinel and Landsat products) as well as improvements in post-processing techniques unlocked through machine learning have accordingly led to increasingly accurate dasymetric population modelling which produces data sets available at a global scale (Nieves et al., 2017). An example of a global population dataset is the Gridded Population of the World Layer (currently GPWv4), which provides 1-kilometer resolution population estimates derived from a proportional allocation gridding algorithm (Doxsey-Whitfield et al., 2015). WorldPop is another example, which produces country-level layers at 100-meter resolution using a weighting scheme based on a random forest, nonparametric predictive model which improves incrementally as more spatial covariates are introduced (Stevens et al., 2015). The High-Resolution Settlement Layer (HRSL) has produced 30-meter resolution population data sets for 18 countries applying an approach for identification of building footprints which uses convoluted neural networks (Tiecke et al., 2017). While each method has its own relative strengths, the important takeaway from progress over the past decade is that the resolution in which we can accurately

predict the distribution of populations spatially has been incrementally improving.

Geostatistical modelling approaches are often applied in fields such as epidemiology (see the Malaria Atlas Project (www.malariaatlas.org) and the Institute for Health Metrics and Evaluation (www.healthdata.org)), and to better understand socioeconomic characteristics of populations in developing countries (see WorldPop (www.worldpop.org)). Alegana et al. (2018) show how geostatistical methods can be effective: a Bayseyian item response theory framework is used to better understand treatment seeking behaviour for children with fever by triangulating household survey data with georeferenced public health facility data. Utazi et al. (2020) also use a geostatistical modelling approach to produce 1-kilometre resolution estimates for coverage of measles vaccine in Nigeria.

While certainly an effective approach, geostatistical methods can be complemented by techniques that generate spatially referenced synthetic populations for usage in agent-based modelling. Several recent efforts have focused on producing spatially referenced synthetic household data. Thomsen et al., (2018) use simPop and ArcGIS to derive a spatially referenced synthetic population in Namibia's Oshikoto region. RTI, an independent, non-profit research institute, has produced a synthetic household population database for 116 million households in the United States (RTI, 2013). The primary contributions of this paper are: (i) to present a novel methodology for the generation of spatially referenced synthetic households by combining spatial microsimulation and dasymetric modelling techniques; and, (ii) to present a case study that includes field validation from a georeferenced survey.

4.2. Methodology

Our proposed methodology combines spatial microsimulation techniques for

generating synthetic populations with dasymetric modelling techniques for determining population distribution to produce a dataset of spatially referenced synthetic households in the form of point data with attached attributes. Our methodology entails three discrete steps: 1) generation of synthetic households inclusive of selected attributes to a certain level of aggregation; 2) spatial disaggregation of household units using dasymetric modelling techniques; and, 3) pairing of synthetic households with location of household units based on the population distribution identified in the dasymetric model.

4.2.1. Defining the source areas and associated household aggregates

We start by defining l non-overlapping source areas $A_i: i = 1, \dots, I$ where

$$\bigcup_{i=1}^{I} A_i = A$$

Each A_i is known to contain n_i households. Source areas would typically pertain to discrete administrative units (e.g. districts, states, etc.), and the number of households within each source area would typically be informed by a population census where data aggregates are made publicly available as attributes linked to those discrete administrative units.

4.2.2. Generate synthetic households

The methodology we use for the population synthesis uses the framework provided by the R package simPop (Templ et al., 2017) which is an open-source and flexible synthesiser based on a modular object-oriented concept. It relies on aggregate tables for population margins and a microdata sample, both ideally from a population

census, where the microdata sample contain at least all variables available in the aggregated tables. The steps for generating the synthetic population can be summarised as follows:

- Calibrate survey weights in the microdata to fit population margins
- Extrapolate the microdata using alias sampling to generate the synthetic population
- Model variables of interest, household or personal variables, on the microdata and "predict" them for each person in the synthetic population
- Use simulated annealing (SA) to the synthetic population to known population margins

SA is the heuristic procedure used to approximate complex optimisation problems (Templ et al., 2017). The SA algorithm has shown to outperform other methods, including deterministic re-weighting and conditional probabilities (Harland et al. 2012).

4.2.3. Calibration

If the available microdata does not already have suitable sampling weights they need to be created or fitted to known population margins. Ideally one should use the available aggregate tables and constructing suitable sampling weights. For the calibration technique we use iterative proportional fitting (IPF) (Deming and Stephan 1940), also known as matrix raking.

4.2.4. Extrapolate microdata

Next, we define the household structure of the synthetic population, using variables

such as age, sex, geographic areas and combinations thereof. Variable for geographic areas must at least include the non-overlapping source areas $A_i: i = 1, \dots, I$ or a more detailed sub-grouping of those. With these basic variables, including a personal and household identifier, the structure of the synthetic population is simulated using the calibrated survey weight and alias sampling (Walker, 1977). The resulting synthetic population contains each of the variables selected for defining the household structure. Additional variables can be added to the synthetic population using modelling techniques.

4.2.5. Model variables

The use of synthetic population greatly increases with having a richer set of variables available for each unit. To enrich our synthetic population, one can use models fitted to the survey microdata to predict additional variables onto each unit in the synthetic population using. Those can be personal variables as well as household variables. Let $\mathbf{v} = (v_1, v_2, \dots, v_p)$ be a vector of variables available for some or all units in the microdata as well as the synthetic population and y_j a variable available in the microdata. We fit a model f depending on the set of variables in **v** to all units j in the microdata where y_j is non-missing:

$$y_j = f((v_1, v_2, \dots, v_p)_j, \theta) + \epsilon \quad \forall j: y_j \text{ non} - \text{missing}$$

with ϵ as a random error term and θ as a set of parameters which are estimated by fitting the model. The variable y is then predicted for each unit i from the synthetic population using $\hat{y}_i$

$$\hat{y}_i = f((v_1, v_2, \dots, v_p)_i, \theta) \quad .$$

The choice of model mainly depends on the type of variable, for instance logistic

regression for binary variables or linear regression for continuous models. The choice of dependent variables should be chosen based on expert knowledge and tested using out of sample prediction techniques like k-fold cross-validation.

The synthetic population now holds $u_i, i = 1, \dots, N_u$ people within N_h households $h_j,\ j =, \dots, N_h$. Each person has a set of variables $\mathbf{x} = (x_1, \dots, x_{\tilde{q}})$ which have either been simulated through the household structure or predicted using a fitted model. Population margins derived from this synthetic data set might deviate from known population aggregates. This can easily be the case for any of the population margins derived using a predicted variable since the underlying models are simplifications of the underlying true mechanisms and they also include random error terms. Thus, the synthetic population needs to be adjusted to fit specified population margins. The main idea of the procedure is to duplicate the synthetic population k times, then choose a sub-population thereof which fulfils each of the population margins up to a pre-specified error ϵ. Since this is a combinatorial optimisation problem, we use simulated annealing, as implemented in R package simPop, to solve it. The simulated annealing procedure starts by selecting each of the units in the original synthetic population and the swapping households at random between the set of already selected households and non-selected households, e.g. the $k - 1$ duplicates of the synthetic population. The procedure stops if the margins of the selected units satisfy the supplied population margins or the number of iterations or time elapsed reaches a maximum.

4.2.6. Applying dasymetric mapping on the synthetic population

The synthetic population now holds $h_i, i = 1, \dots, n_i$ households in each source area

A_i. Each unit u_j in each household has a vector of attributes $\mathbf{v} = (v_1, v_2, \dots, v_q)$ which can be associated either with the entire household (e.g. source area A_i of household, house structure type, etc.) or to individuals within the household (e.g. age, sex, etc.). The aim of dasymetric mapping is to enhance the geographic resolution of the synthetic population, e.g. mapping each household first to higher resolution grid cells, or eventually to unique x and y coordinates.

4.2.7. Defining sets of spatially disaggregated synthetic households within each source area into smaller target cells

For dasymetric population modelling, we first partition the source area A into many smaller target cells of equal area $C_j: j = 1, \dots, m$. Each cell C_j is then allocated to one and only one source area A_i. There will inherently be cells which intersect the boundary of a source area to which the modeler must define a rule for determining singular allocation. For example, one could assign a target cell C_j to the source area A_i if the centroid of C_j, γ_j lies in A_i, $\gamma_j \in A_i$.

Thus, each A_i is is composed of m_i cells which we label $C_{i,j}: j = 1, \dots, m_i$. We then define a density variable $d_{i,j}$, with:

$$d_{i,j} \geq 0 \quad \wedge \quad \sum_{j=1}^{m_i} d_{i,j} = 1$$

The density variable may be obtained by combining a binary masking variable $b_{i,j}: b_{i,j} \in \{0,1\}$ with some other variables $z_{i,j}$ which contain allocated values of a specific attribute for each $C_{i,j}$. A binary variable should indicate whether people are living in a grid cell, such as the existence of housing infrastructure. For example, if

housing infrastructure exists in target cell $C_{i,j}$ then $b_{i,j} = 1$, if not then $b_{i,j} = 0$.

The density variable $d_{i,j}$ can be calculated through

$$d_{i,j} = \frac{b_{i,j} z_{i,j}}{\sum_{j=1}^{m_i} b_{i,j}\, z_{i,j}}$$

The expected number of households, $\overline{N}_{i,j}$, in target cell $C_{i,j}$ is then defined by

$$\overline{n}_{i,j} = d_{i,j} n_i \quad .$$

Since $\overline{N}_{i,j}$ is not necessarily an integer value we define an integer realisation of the number of households in $C_{i,j}$ by

$$\hat{n}_{i,j} = \lfloor \overline{n}_{i,j} \rfloor + \delta$$

where $\lfloor x \rfloor$ denotes the largest integer smaller or equal x and $\delta_{i,j}$ a random variable taking the values 0 or 1.

Within each target area A_i the realisations of the sum of random variables $\delta_{i,j}, j = 1, \dots, m_i$ must equal $n_i - \sum_{j=1}^{m_i} \lfloor \overline{n}_{i,j} \rfloor$ thus

$$\delta_{i,j} = \begin{cases} 1 & \text{if} \quad x_{i,j} \leq x_{(t)} \\ 0 & \text{otherwise} \end{cases}$$

$$t = n_i - \sum_{j=1}^{m_i} \lfloor \overline{n}_{i,j} \rfloor$$

with $(x_{i,j}, \dots, x_{i,m_i})$ a vector of random values drawn from a uniform distribution $U(0,1)$ and $(x_{(1)}, \dots, x_{(m_i)})$ the order statistic of these randomly generated values.

4.2.8. Adding coordinates to synthetic households

The synthetic households $h_i, i = 1, \dots, N_h$ can then be disaggregated to an individual target cell $C_{i,j}$ thus creating spatially-referenced synthetic households $\hat{h}_{i,j}, i =$

$1, \ldots, n_{i,j}$. This can be done randomly if spatial dependencies of attributes are insignificant or not considered for the analysis. The random selections must consider that x and y coordinates of households may not overlap. Using random selection can be justified easier in cases where the target cells $C_{i,j}$ cover relatively small areas (e.g. less than 100 square meters). If specific attributes show spatial dependencies, hybrid techniques can be deployed which integrate spatial covariates to weigh the spatial distribution of the synthetic households h_i (Monteiro et al., 2017). The next step entails creating point values for the spatially referenced synthetic households $\hat{h}_{i,j}$ within each target cell $C_{i,j}$. The precise x and y coordinates of the point values for spatially-referenced synthetic households $\hat{h}_{i,j}$ can be generated randomly within the target cell $C_{i,j}$ or can be weighted or snapped to a spatial covariate (e.g. building footprint centroids). The new spatially-referenced synthetic households $\hat{h}_{i,j}$ are assigned spatial attributes x and y in addition to those joined from the synthetic household h_i such that for each $\hat{h}_{i,j}$ holds variable vector $\mathbf{v} = (v_1, v_2, \ldots, v_q, x, y)$.

4.3. Case study: Dumuria Upazila, Bangladesh

To demonstrate the efficacy of these methods, we look at three unions in Bangladesh: 1) Bhandar Para; 2) Sahas; and, 3) Sarappur. All three are located within Dumuria Upazila, Khulna District. Taken together, the three unions are contiguous, rural and located primarily within Polder 29 in the Bangladesh Coastal Zone.

In total, our study area is 83.96 square kilometres and contains 51,497 residents within 12,697 households. We define our study area as all areas covered by the 2011 National Census polygons for our three selected unions (administrative level 4). A summary of key attributes for each union can be found in Table 12 and a map of

our study area can be found in Fig 11. Our three unions contain a combined 71 *mauzas* (administrative level 5). The average *mauza* in the study area is 1.18 square kilometres. However, there is a large variation in size, with the largest being 6.69 square kilometres and the smallest 0.15 square kilometres. The 2011 National Census only covers "populated *mauzas*", of which 59 *mauzas* within the study area are classified.

Table 12. *Area, Population, and Household Data for Bhandar Para, Sahas, and Sarappur Unions*

	Area (sq. km)	**Population**	**Households**
Total	***83.96***	***51,497***	***12,697***
Bhandar Para	34.89	16,749	4,077
Sahas	25.29	18,647	4,498
Sarappur	23.78	16,101	4,122

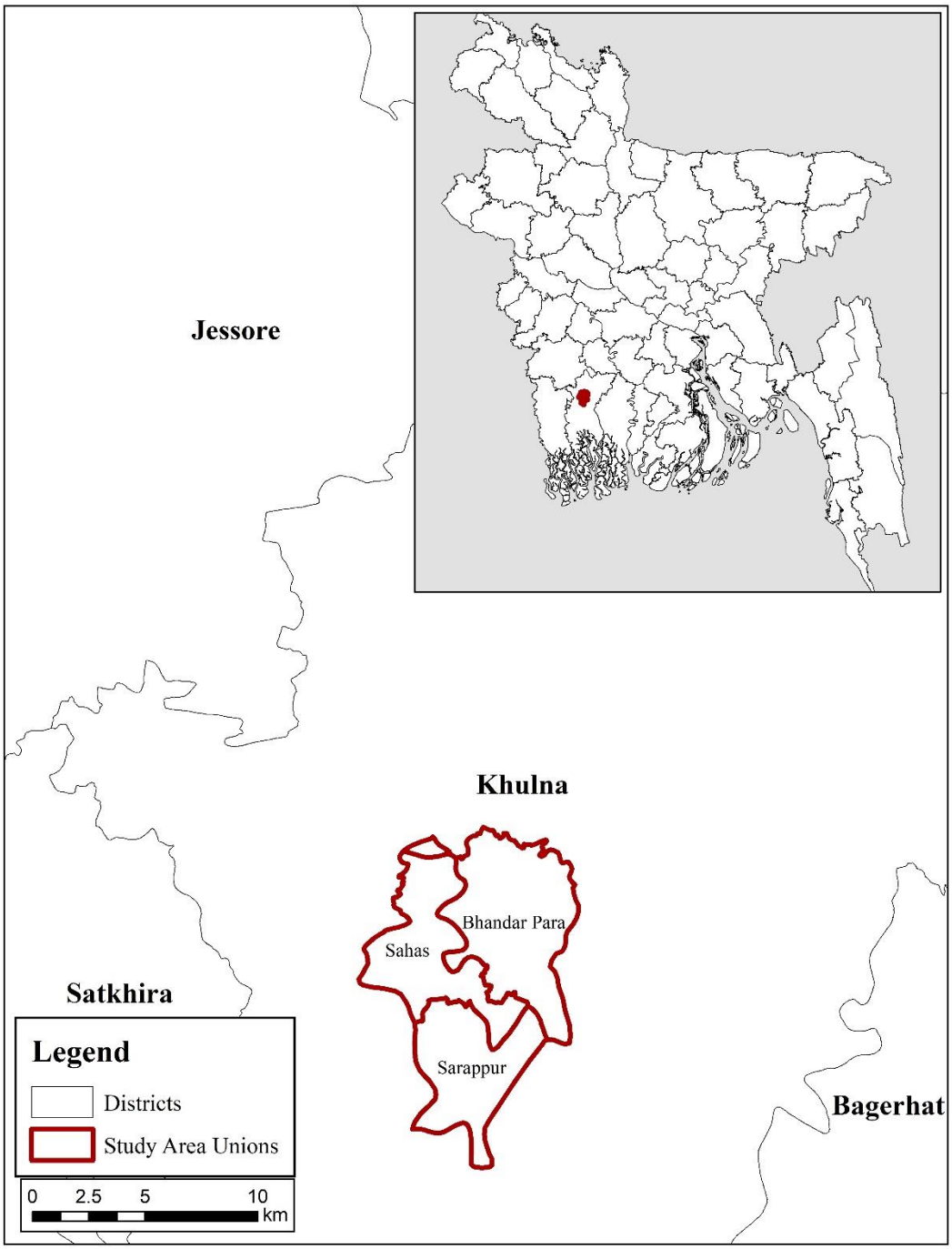


Fig. 11 Map of Study Area Unions

Our study area was selected in consultation with the REACH Programme, a global research programme aimed at improving water security for the poor. The REACH Programme has a sub-project within Polder 29 which includes a georeferenced household survey. Polder 29 provides an interesting case for further

usage of our output dataset for agent-based modelling. Polders can be modelled as enclosed systems within the context of broader biophysical conditions in which the interactions between flood events, salinity, water infrastructure, agricultural production and income can be tested (Borgomeo et al., 2017). Each of these inherently has a spatial element within a polder system, and the output dataset from this study may provide more insight into those spatial heterogeneities. Taken together, our study area presents an opportunity to apply our methodology in a well-mapped and uniquely studied coastal system which provides the means for on-the-ground verification of results.

4.3.1. Data

For census aggregates, our study utilizes data from the 2011 National Census, produced by the Bangladesh Bureau of Statistics (BBS). Census data were used both in the dasymetric mapping segment and the population synthesis segment of our study. For dasymetric mapping, the union polygons provided the basis for spatial disaggregation of households, while the Mauza polygons provides a means for verification through reaggregation of households into lower-level administrative boundaries. For population synthesis, microdata from the 2011 National Census released through the Integrated Public Use Microdata Series (IPUMS) for Dumuria Upazila were used to extrapolate household composition and attributes.

Our study area was also the target of a “mapathon”, which took place over a period of five hours on August 25, 2017 by volunteers mobilised from OpenStreetMap (OSM) Bangladesh. Specific features mapped included over 22,000 building footprints, 200 roads, 100 waterways, and 300 land-use polygons. These

features, particularly building footprints, are important for the dasymetric population modelling section of our study, which uses the layer as a binary mask for determining areas of potential habitation.

Lastly, this study relies on a first phase survey by conducted by the REACH Programme in Polder 29 from December 17, 2017 to January 7, 2018. The survey collected attributes for 707 households within our study area, representing 5.6 percent of all households. For the survey, households were selected within each *mauza* to ensure a representative distribution, and GPS coordinates were logged of each household location. Individual household locations were used to validate the output dataset from this study.

A summary of data layers used in the analysis can be found in Table 13.

Table 13. *Data Layers Used in Dumuria Upazila Analysis*

Type	Description	Source	Pop Synth	Pop Mod	Field Ver
National Census Aggregates/ Shapefiles	2011, administrative levels 3 (Upazila), 4 (Union) and 5 (Mauza)	Bangladesh Bureau of Statistics (BBS)			
National Census Microdata	2011, administrative level 3 (Upazila)	BBS			
Building Footprints	2017, manually traced	OpenStreetMap (OSM)			
Polder 29 REACH Survey	2017 – 2018, 707 households within study area	REACH Survey			

4.3.2. Analysis and Results

Population Synthesis

To illustrate the functionality of our model for generating spatially referenced synthetic households, we rely on the R package simPop for population synthesis. The workflow used can be viewed in Fig. 12.

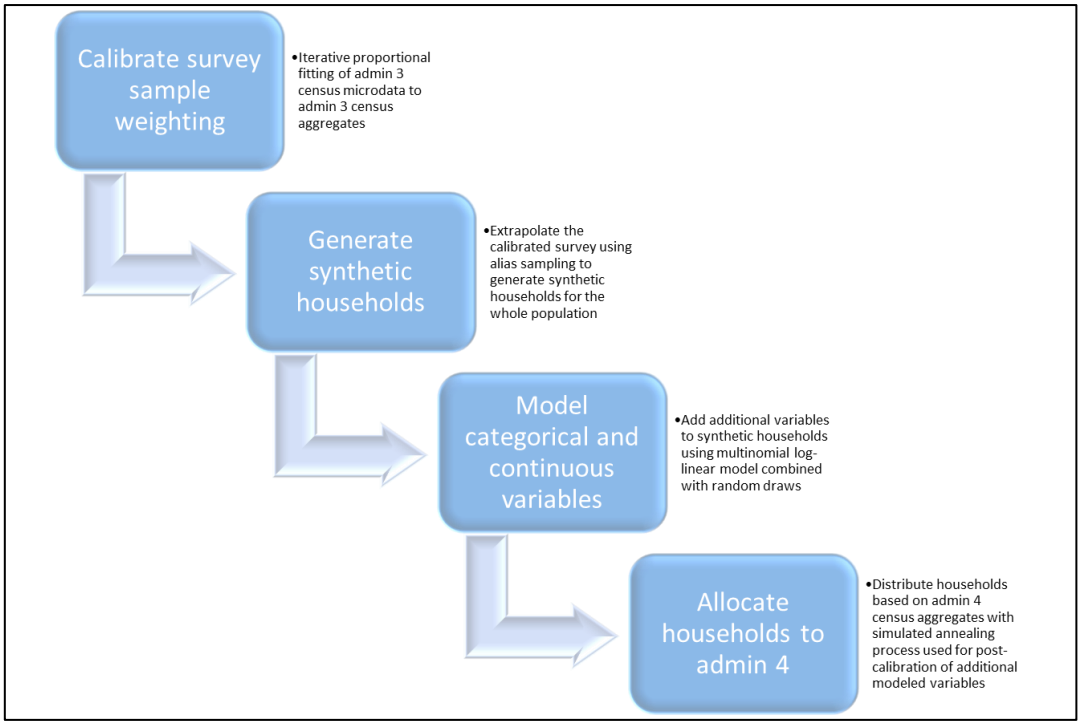


Fig. 12 Workflow for Simulation of the Population with SimPop

We start by calibrating the sample survey microdata for Dumuria Upazila to its census aggregate data using iterative proportional fitting, implemented in the R-package surveyed. Next, we define the household structure of the synthetic population by extrapolating the calibrated survey sample using alias sampling (Walker, 1977) with attributes for household and personal identifiers, household size and sampling weight. Because Dumuria is the only *upazila* considered in our analysis we do not need to consider strata. Next, we separately model continuous and categorical variables of individuals within each household. The only continuous variable we consider in our analysis is age. For categorical variables, we consider sex (male or female) and relation to head of household (spouse_partner, child, other_relative, or non_relative). For both, we use simPop's built-in functions to employ a multinomial log-linear model combined with random draws (Templ et al., 2017).

We next employ simPop's function for allocating households into smaller areas, which allows us to place our synthetic households into discrete unions as target areas $A_i: i = 1, \ldots, I$ within Dumuria Upazila. This process uses a simulated annealing algorithm, which entailed an iterative process of swapping households for optimal combinations based on our known margin of total population within each union. Our final output is synthetic population data for all unions within Dumuria Upazila, although only three are considered further in our analysis. Selected combined results for the three unions considered can be viewed in Figs. 13 and 14.

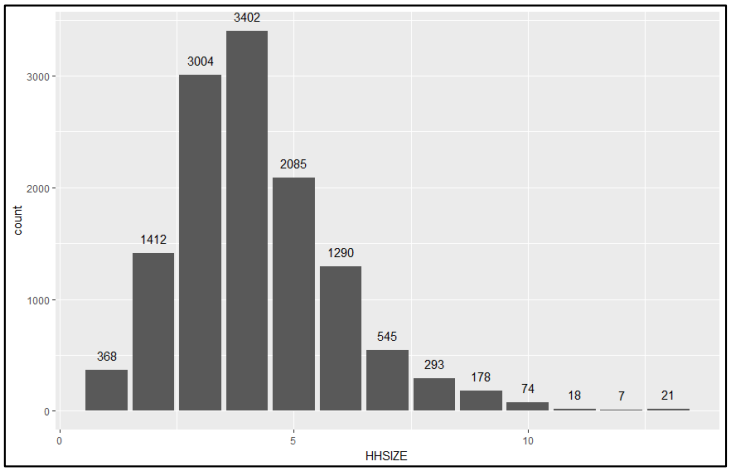


Fig. 13 Combined distribution of household size within study area

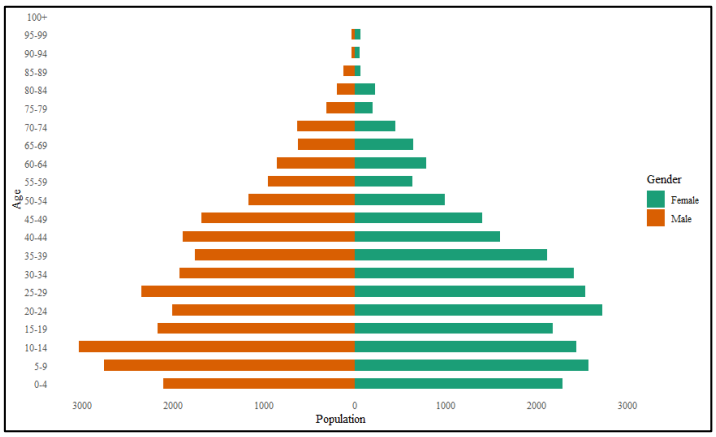


Fig. 14 Combined synthetic population pyramid for study area

Dasymetric Population Modelling

For dasymetric population modelling, we start by applying a 30-metre resolution grid across the study area. We selected 30-metre resolution to align with the resolution presented by the HRSL. We then overlay building footprint polygons and calculate the area within each cell covered by a building footprint polygon. The total number of households within each union (our target area A_i) are then allocated to cells $C_{i,j}$ within that union using the building footprint polygons as a binary mask and the building footprint area within each cell as a proxy for density. To calculate our density variable, we divide the footprint area within each cell by the sum of building footprint areas within each union and multiply by the population of the union to get a real number of households per cell $\hat{n}_{i,j}$. From there, we separate the integer from the decimal part and classify all integers as households. Decimal parts are then used for weighted random allocation of remaining unallocated households within each union. The resulting layer is of 30-meter resolution gridded households, where each cell $C_{i,j}$ is associated with an attribute for $\hat{n}_{i,j}$households. Lastly, $\hat{n}_{i,j}$ random points were created within each cell $C_{i,j}$ using the Create Random Points tool in ArcMap 10.6. The process for achieving modelled household points can be seen in Fig. 15.

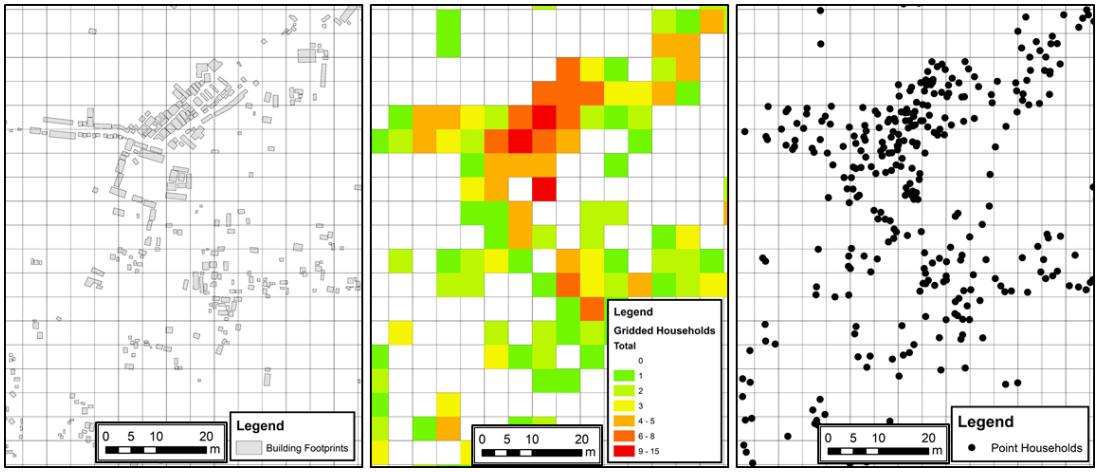


Fig. 15 30-meter resolution grid cells overlain by building footprints (left), 30-meter resolution modelled gridded households (centre), and modelled household points (right) in Sahas Union, Khulna District

Spatially Referenced Synthetic Households

The last step entails joining the tabular synthetic households with the point features. We first randomise the order of the synthetic households then assign each a unique identification attribute. We then execute a join (table to layer) which matches each point with a synthetic household by matching the unique identification attributes of both. This is done for each of the three unions. After which, each synthetic household point feature contains attributes such that $\mathbf{v} = (v_1, v_2, \ldots, v_q, x, y)$. The results can be seen in Fig. 16.

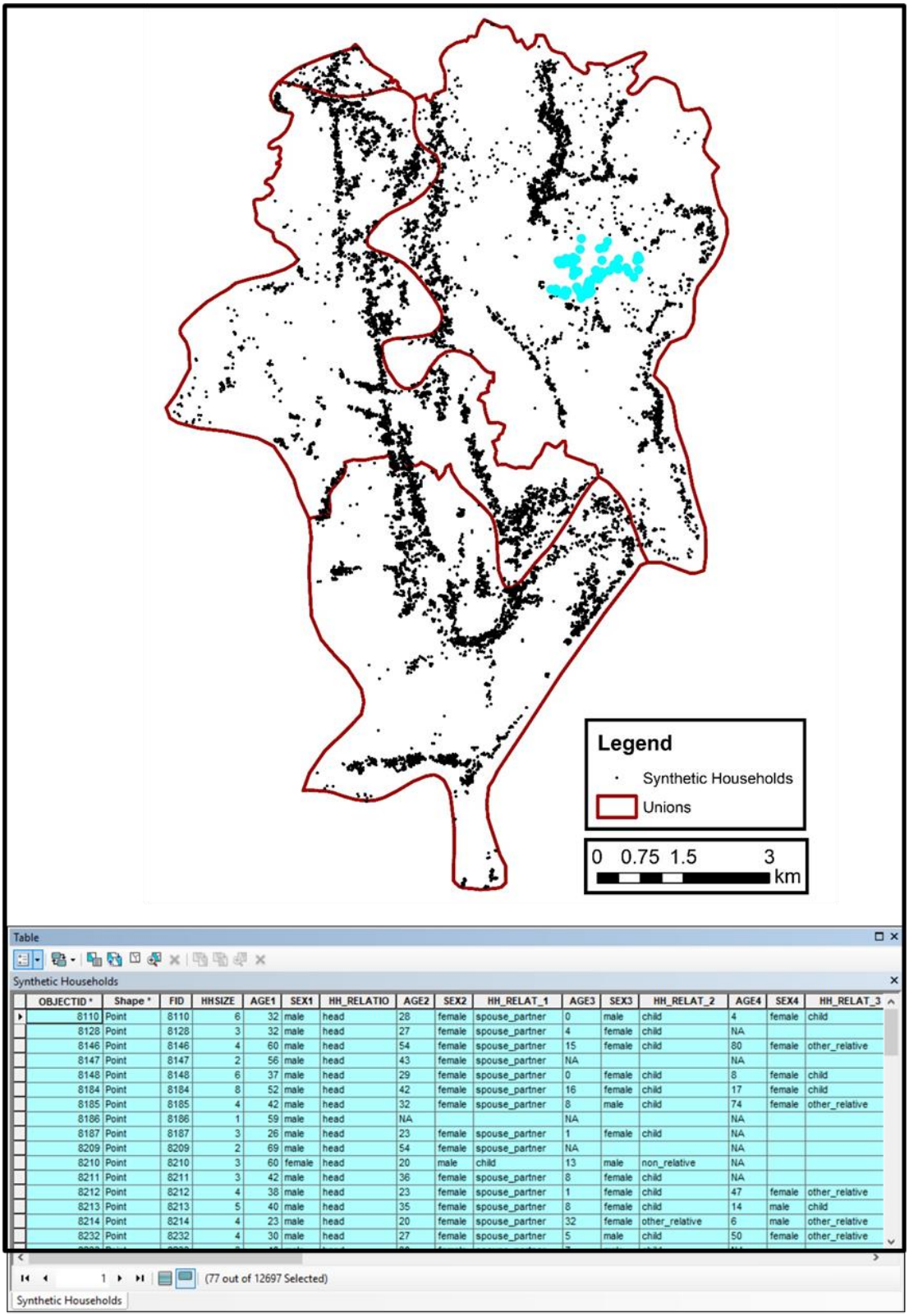


Fig. 16 Point features for spatially referenced synthetic households and sample from accompanying attribute table for selected features (in light blue)

4.3.3. Validation

Field validation of household locations through the REACH Survey

To assess the performance of our results, we benefit from access to the spatially referenced REACH survey. The REACH survey contains 707 households comprised of 3,124 persons within our study area in comparison to our 12,697 synthetic households comprised of 25,964 persons. We first compare the age distribution of the surveyed population with that of our synthetic population. Although our synthetic population provided real numbers 0 – 99, the survey only provided categorical information which we used to aggregate our synthetic populations. Although the synthetic populations were skewed towards a younger population, results in all categories are within five percent of the surveyed population. The results can be seen in Fig. 17.

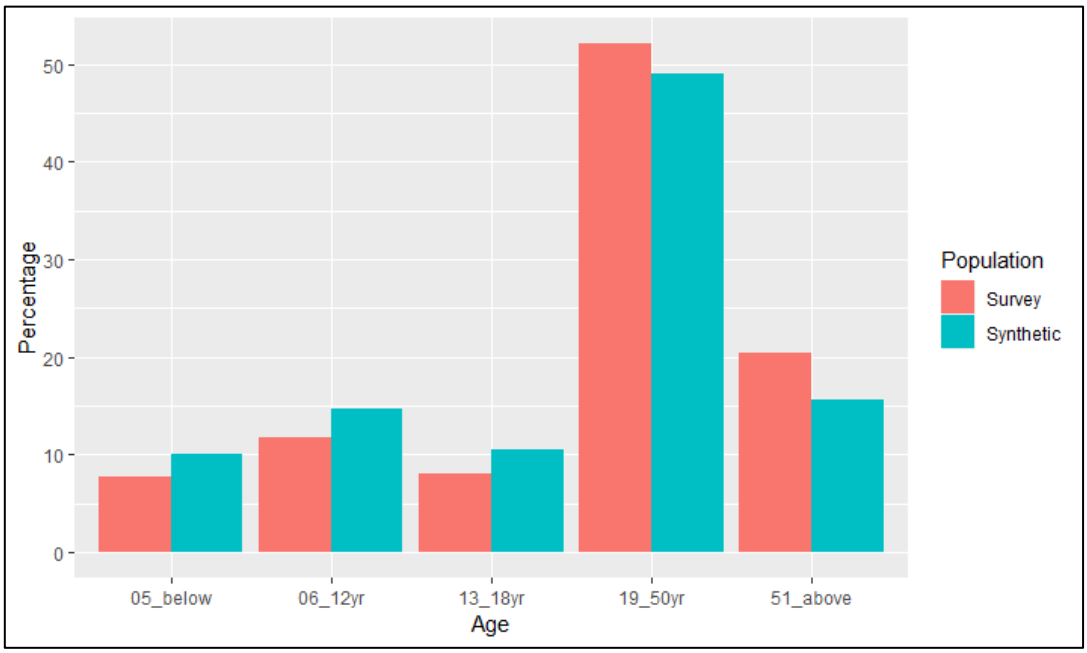


Fig. 17 Age distribution of surveyed vs synthetic populations

We then use the geolocation of the households surveyed to validate the location of our synthetic households. We do this by calculating for each surveyed household the distance to the nearest synthetic household. Although the analysis reveals a few distinct outliers, 98 percent of all surveyed households were within 100 meters of at least one synthetic household, and the median distance observed was only 18 meters. This pattern can be clearly detected in Fig. 18.

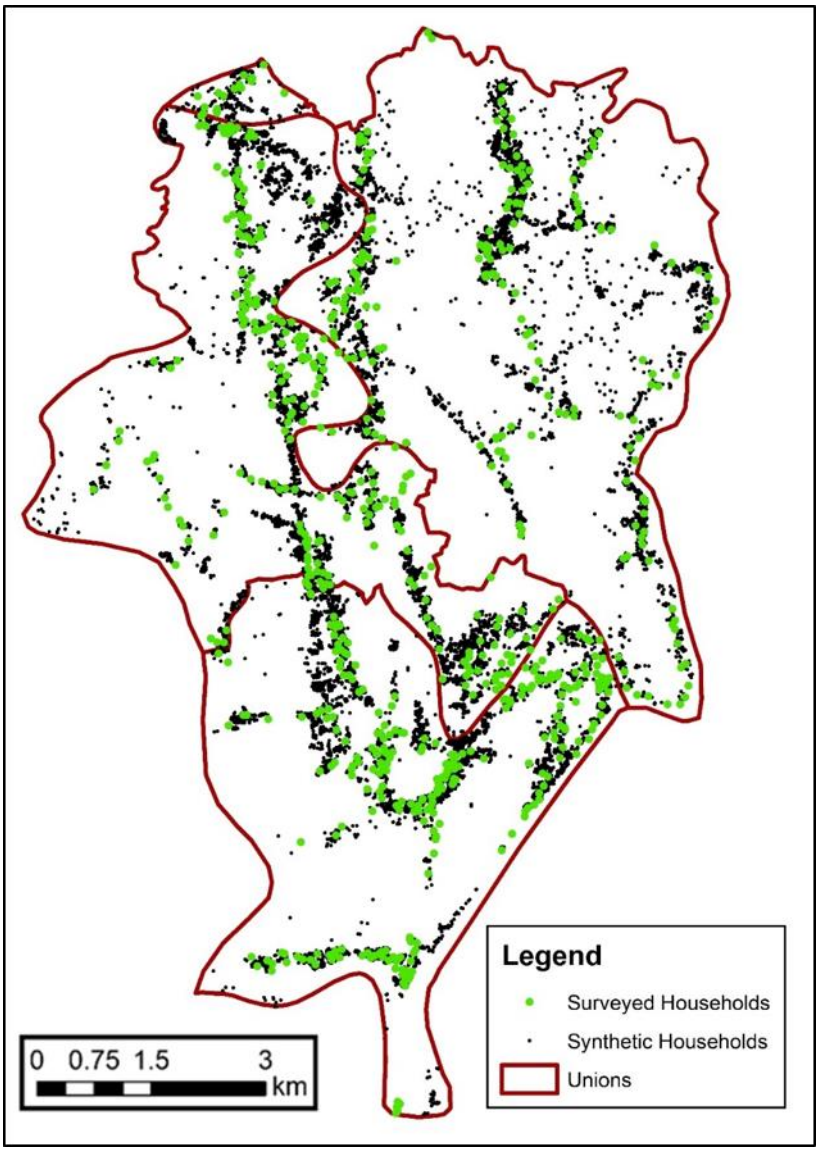


Fig. 18 Point features for spatially referenced synthetic households overlaid by point features for surveyed households

Performance through RMSE and MAE of lower-level administrative aggregates

Our model spatially disaggregates synthetic households at the union level. To validate our results, we reaggregate our synthetic households (households predicted) at the *mauza* level and compare results to the census aggregates (households observed). However, we only do this for the 27 *mauzas* within Sahas and Sarappur Unions as there are noticeable errors in the census aggregates for Bhandar Para Union visible in Fig. 19. Sitarampur Mauza in particular shows clear patterns of settlement despite being classified as unpopulated per the 2011 National Census. These errors were not observed within *mauzas* of Sahas and Sarappur Unions.

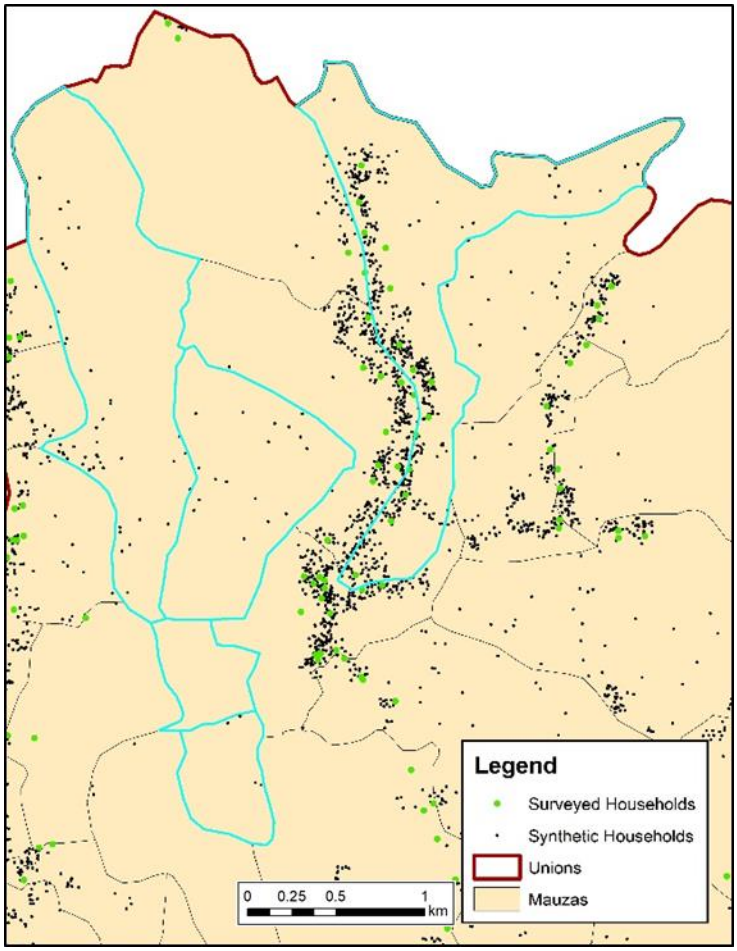


Fig. 19 Selected *mauzas* within Bhandar Para Union are designated as "NF BBS 2011" per the 2011 National Census and are listed as unpopulated

The performance of our model can be seen in Fig. 20. We generate the root mean square error (RMSE) and mean absolute error (MAE) of lower-level administrative aggregates. We also express each of the errors as a percentage of the average number of households per *mauza* tested.

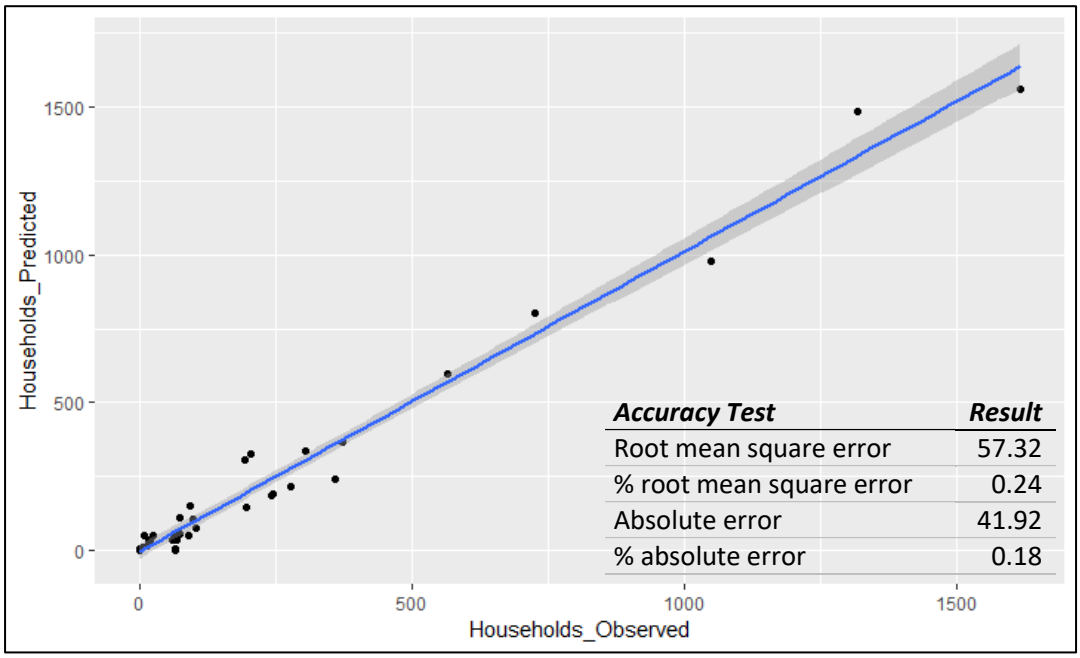


Fig. 20 Linear model of households observed vs households predicted with 95 percent confidence interval and results of accuracy tests for 27 *mauzas* within Sahas and Sarappur Unions

4.4. Discussion

4.4.1. Introducing a Paired Approach for Generating Geospatially Referenced Synthetic Household Agents

The methodology proposed in this paper effectively combines spatial microsimulation techniques for generating synthetic populations with dasymetric modelling techniques

for determining the spatial distribution of populations to produce a dataset of spatially referenced synthetic households in the form of point data with attached attributes. This represents the natural convergence between two prominent geospatial workarounds for better understanding population dynamics without having full access to primary information. While this technique is exceedingly difficult to verify at lower resolutions (perhaps why it has yet to be done), our study has the advantage of being able to draw upon a spatially-referenced survey of households to verify both modelled location information and associated household attributes in a contained geographic area.

Although the circumstances at first appear unique, they will become increasingly accessible with advancements in technology and proliferation of publicly available microdata and geospatially referenced survey data. The techniques used for high-resolution dasymetric mapping in particular are increasingly achievable at higher accuracy and at a larger scale. At the same time, the increased prevalence of higher resolution microdata releases and advancements in spatial microsimulation techniques have enhanced the ability in which we can confidently predict socioeconomic characteristics of those mapped populations. Efforts such as IPUMS and other data repositories have facilitated this by encouraging the open release of microdata and cataloguing its availability for public usage. Still, public releases of census microdata only cover a small percentage of the total individuals and households for the aggregated unit (typically up to 10 percent), thus making population synthesis necessary to generate attributes for the entire population reflective of its general composition.

While SAE techniques can increase the resolution in which we understand the

spatial distribution of attributes, they are still restricted to aggregates so would not pick up on clusters and patterns of settlement within those aggregated areas. Thus, despite an inability to effectively differentiate household attributes spatially within the finest level of aggregation, it is still a worthwhile exercise to disaggregate those households in line with spatial patterns of development. This is particularly true if only interested in attributes which would not have strong spatial dependencies within a small administrative unit, such as household size within predominantly rural areas, or composition of sexes within a household. For spatial attributes with stronger potential spatial dependencies (e.g. wealth, water access, etc.) the spatial distribution can be further refined within the administrative unit by incorporating spatial modelling techniques which consider factors such as the spatial autocorrelation of selected attributes, dependencies related to additional known covariates (e.g. poorer households may live further from water access) or additional georeferenced surveys which can unlock spatial dependencies for attributes collected which are also contained within the census data. When analysis is combined with other spatial data sets (e.g. flood risk, groundwater contamination, etc.), even a limited number of attributes can prove quite useful when households are fixed to a geographic location.

While recent efforts have been made to better approach challenges of geographic partitioning for data anonymisation, none have resolved the fundamental issue that providing a spatial reference compromises the anonymity of potentially sensitive data (Croft et al., 2017). Certain applications, particularly spatial ABM modelling and spatial microsimulation stand to benefit heavily from household-level spatial information. For these applications, spatially referenced synthetic populations could provide a workable solution for preserving security of potentially sensitive data. Spatial ABMs require a modelled landscape to capture spatial heterogeneities of

inputs, which our proposed method can provide. These modelled landscapes are often comprised dynamic networks of interacting agents, such as in an ecosystem, city, or financial market (Grimm et al., 2006). The challenge in spatial ABMs is to capture the interplay between discrete representations of space and network or diffusion models, so the more detail available on the discrete representations the better. Past models have done this through aggregated administrative zones (Smajgl and Bohensky, 2013), division of "city area types" (Gaube et al., 2013), high-resolution land use (Sun and Muller, 2013), and various other criteria. The approach that we propose in this paper advances the resolution dramatically, by specifying the precise location of each household within the synthetic population set through a novel method which incorporates the latest techniques of dasymetric population modelling. Using our proposed method, discrete representations could thus include individual households.

The primary advantages of distributing the populations spatially into discrete households are two-fold: 1) it allows for spatial patterns of settlement to be considered for spatial analysis, effectively breaking down barriers of potentially arbitrary administrative units; and, 2) it allows for spatial ABM of populations into complex ecosystems, including the interaction with other mapped physical features including roads, rivers, economic centres, etc. The outputs of our study provide a ground-truthed first take at producing spatially referenced synthetic households using only census aggregates, household microdata, and gridded population data; all of which are readily available for most countries in the world.

4.4.2. Applying our model to enhance local decision-making capabilities

The output dataset of spatially referenced synthetic households can be used as an input for further spatial analysis and agent-based modelling. An example of this at a basic level in the Bangladesh Coastal Zone would be to use a roads layer to calculate travel times of households to cyclone shelters. This would be helpful for prioritisation of new cyclone shelter construction, as well as planning of outlet roads. Similar accessibility layers could be produced for access to schools or health centres.

Flood risk is another domain where our output datasets may provide utility. To exemplify this, we overlaid our spatially referenced synthetic households layer with a layer showing areas that have been inundated at least once from 1985-2010 (Winsemius et al., 2013) (Fig. 21). We found that roughly 15 percent of households lie in a flood prone area despite 36 percent of the study area lying in a flood prone area. It makes conventional sense that human settlement would avoid flood prone areas; however, with only administrative boundaries and census aggregates it is difficult to come to a substantiated conclusion. We can take this a step further to analyse information derived from household and individual attributes. From our output layer, we can show that approximately 3,000 children live within flood prone areas, with children represented in 86 percent of all exposed households.

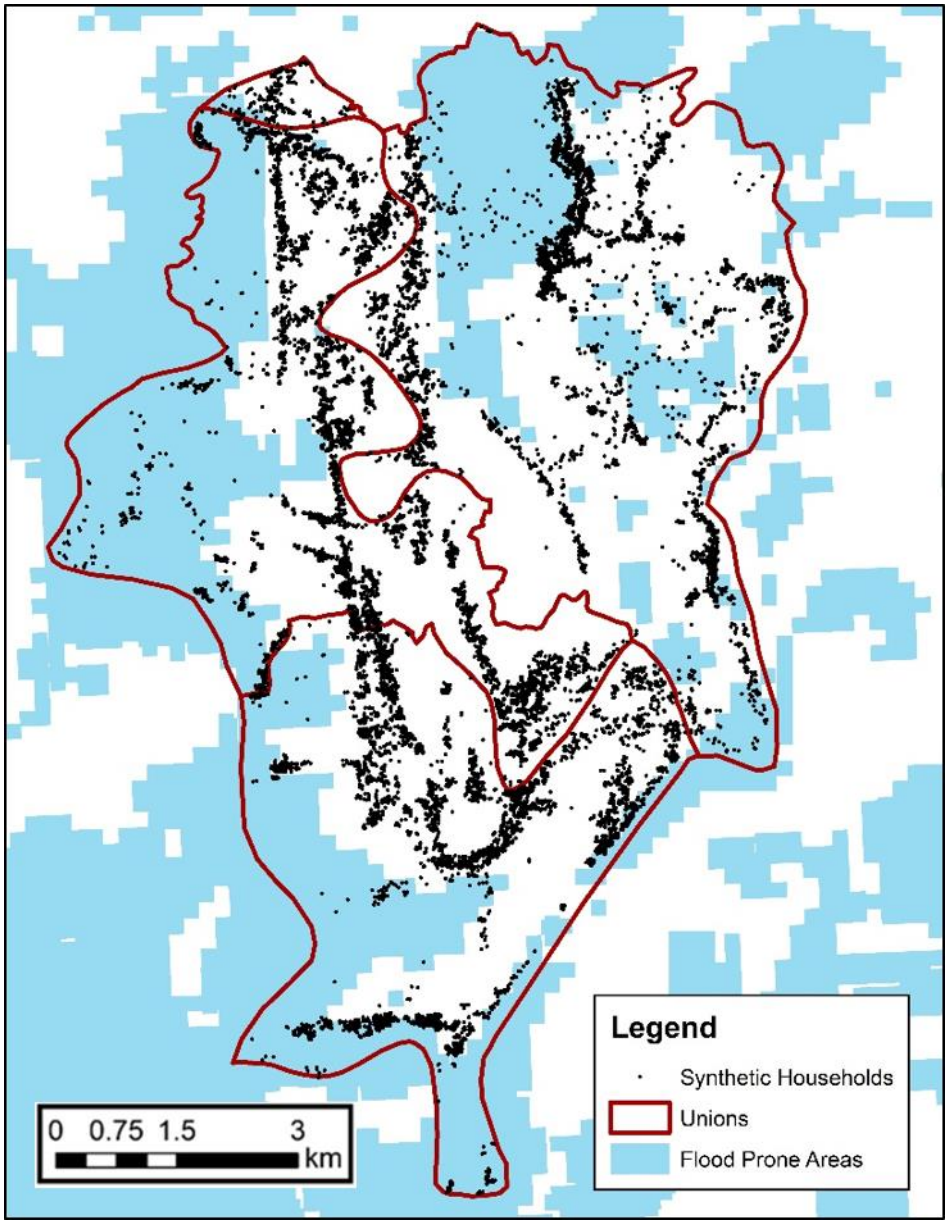


Fig. 21 Overlay of areas that have been inundated at least once on the full Dartmouth Flood Observatory record (1985-2010) and spatially referenced synthetic households.

This information can be helpful to improve analysis of biophysical characteristics of human settlement patterns and their corresponding socioeconomic attributes. Agent-based modelling can then be employed to analyse the movements and effects of flood events on household agents. Each household can thus be viewed as either a fixed automaton or a non-fixed automaton depending on the attributes

considered (e.g. house structure is a spatially fixed asset whereas value attached to the individual household members would be consider non-fixed). This would be helpful for analysing the impacts on welfare of households affected by disasters.

4.4.3. Data-driven advancements

Our model is designed to be flexible to advancements in technology on multiple fronts. Dasymetric population modelling depends heavily on earth observation derived data. Increasingly, this data is becoming available at higher-resolutions and at lower costs. Free satellite data from the European Union's Copernicus Programme combined with better post-processing techniques which rely heavily on machine learning have led to more and better information at scale. Machine learning techniques will also help to close the gap on modelling socioeconomic characteristics. At a more localised level, the use of drones has proliferated to enable quicker and better access to high-resolution earth observation data including lidar where 3D representations and detailed classification of features can help to further refine physical characteristics of structures which may be indicative of the socioeconomic characteristics of those households living within. This can be supplemented using street view imagery to further classify structural attributes of features such as building structure type and façade, or even associating the brands and models of cars within a neighbourhood to household income (Gebru et al., 2017).

Our study utilises survey data for verification purposes. However, another area of advancement lies in the application of open data such as OpenStreetMap, street view imagery and other surveyed data as training sets for machine learning (Diou et al., 2018). Our model allows for expansion into this area as it provides a top-down

then bottom-up approach. This is particularly true for spatially dependent socioeconomic characteristics where we may have a good sense of the statistical distribution of attributes at an aggregate level but require further spatial information to distribute synthetic populations appropriately. Our study avoids confronting this challenge by only focusing on less spatially dependent attributes such as household size, gender and age composition. However, going forward this is a critical area of growth which should be further considered.

4.4.4. Limitations

In this study, we do not project demographic changes for households listed in the 2011 National Census for 2018 (the year of the REACH survey). However, from the 2001 National Census to the 2011 National Census, the population of Khulna District (administrative level 2) remained relatively steady, decreasing slightly from 2.38 million to 2.32 million (Bangladesh Bureau of Statistics, 2015). This study focuses primarily on validating the precise location of spatially referenced synthetic households, which are unavailable from the 2011 National Census data. Thus, although there may be slight changes from 2011 to 2018 in total population within our study area (e.g. negative population growth due to broader urbanisation trends), we do not see this as a major limitation. The larger limitation is that due to lower fertility rates, the age structure of the population will have shifted: the United Nations estimates that the population aged 0-14 will have reduced from 29.3 percent in 2010 to 26.8 percent in 2020, while the population aged 15-64 will have increased from 65.6 percent in 2010 to 68.0 percent in 2020 (UN DESA, 2017). Furthermore, while the output layer show promise, there is still room to improve the types of attributes included in the analysis by incorporating additional methods for distributing synthetic

population across the study area. This is particularly important for potentially spatially heterogeneous attributes such as household welfare, which may be better informed by including additional spatial covariates such as those related to house structure type, access to water, or type of toilet (see Thomsen et al., 2018), and machine learning methods which use survey data as training sets.

4.5. Conclusions

Our study presents a methodology to generate spatially referenced synthetic households for population modelling. We combine spatial microsimulation techniques to generating synthetic households with dasymetric modelling techniques to inform that spatial distribution of those households. In our case study of Dumuria Upazila, Bangladesh, we highlight the efficacy of our techniques through on-the-ground verification of location and attribute information and re-aggregation into lower-level administrative boundaries. Distributing populations spatially into discrete households provides advantages even where only limited attributes are included. Population is inherently distributed unevenly across a geographic surface, and our model allows for synthetic households to be represented spatially better than through association with potentially arbitrary administrative units. This is not particularly important for ABM that is focused on populations within aggregate areas. However, new branches of ABM which focus on spatial relationships between agents as individual automata can benefit from this data. Household agents can allow for a deeper understanding of the interactions between households and complex geographic systems.

Our method does present some potential challenges and opportunities in terms of privacy considerations. Advancements in spatial modelling and machine learning techniques are enabling us to know much more about geographic surfaces and the socioeconomic characteristics of populations who inhabit them. While this could present a substantial opportunity to improve our insights into everything from poverty reduction programmes to infrastructure decision-making, it also requires careful consideration form the side of citizen privacy rights. It is not unreasonable to expect that one could take a small training set of surveyed household data and extrapolate it to generate accurate and potentially sensitive information related to socioeconomic characteristics such as household income or potential to default on a housing loan. Indeed, some of this may be unavoidable should census micro-data truly be replaced by targeted georeferenced household surveys used as training sets. Our model does not take it that far but does shed light into the future direction of population modelling.

5. High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone

Earth Observation, extended data collection methods and high-resolution modelling are providing more detailed and disaggregated insights for disaster risk management. This paper reviews several of these new data and analysis opportunities and demonstrates their usefulness by integrating them in a cyclone risk analysis in an area of the Bangladesh Coastal Zone that is inhabited by 9.5 million people, using the 2019 Cyclone Fani as an example. The analysis offers new insights into the impact of the cyclone on populations by generating a high-resolution estimate of the flood extent, modelled point features of potentially flooded households, and modelled synthetic structural and socioeconomic characteristics of each household as a proxy for vulnerability. We demonstrate how the analysis supplements locally reported estimates of damaged households by providing a more consistent distribution of flood extent between districts and providing estimates of the relative vulnerability of potentially flooded households across districts. This information can be used in targeting disaster response and recovery, as well as broader efforts to enhance resilience.

Keywords: cyclone; disaster risk management; remote sensing; population modelling; geographic information systems; Bangladesh

5.1. Introduction

5.1.1. Disaster risk reduction as a global priority

The World Economic Forum's *Global Risks Report 2020* lists extreme weather events, natural disasters, and failure of climate change mitigation and adaptation as three of the top 10 perceived global risks, both in terms of likelihood and impact. These concerns appear to be well-grounded considering the increasing prevalence of loss events (730 in 2017 compared to an average of 605 over the last 10 years and 490 over the last 30 years) and increasing overall losses resulting from those events ($340 billion in 2017 compared to an average of $170 billion over the last 10 years) (Munich Re, 2018). Some of this growth is only natural and expected; as countries develop, so too does the value of their potentially exposed assets (Pielke, 2019). However, the increasing exposure may be exacerbated by the absence (or lack of enforcement) of land zoning and other policies to minimise the build-up risk. Some of this increase in risk is also likely due to an increasing frequency and severity of extreme weather events resulting from climate change (Hayhoe et al., 2018).

Disaster risk is defined as the "the potential loss of life, injury, or destroyed or damaged assets which could occur to a system, society or a community in a specific period of time" (UN General Assembly, 2016) and is often described as a function of hazard, exposure, and vulnerability; a concept known as the "Crichton Triangle" (Crichton, 1999). The United Nations has adopted the *Sendai Framework for Disaster Risk Reduction 2015-2030* (UN General Assembly, 2015) to facilitate "the substantial reduction of disaster risk and losses in lives, livelihoods and health and in the economic, physical, social, cultural and environmental assets of persons, businesses, communities and countries".

For countries to achieve the ambitious targets outlined in the *Sendai Framework for Disaster Risk Reduction 2015-2030*, four priorities for action are outlined, the first of which is "understanding disaster risk". The idea behind this priority is simple: better data leads to better information, which enables more informed decision-making across the four phases of the disaster risk management cycle: (i) response; (ii) recovery; (iii) mitigation; and, (iv) preparedness (Khan et al., 2008; Othman et al., 2014). Thus, robust data are critical for effectively coordinating disaster response, identifying recovery needs, informing risk mitigation investment planning, and ensuring preparedness capacity. Fortunately, there is now a proliferation of new data sources, which are enabling disaster risks to be assessed at ever-higher spatial resolutions, providing the basis for more targeted disaster risk management. This paper implements a series of developments in geospatial analysis that together represent a step-change in disaster risk mapping capabilities.

5.1.2. Understanding the spatial characteristics of disaster risk

Continued advancements in high-resolution data collection and modelling techniques, as well as the proliferation of primary and derived data sets for public use, has led to an expansion of our understanding of the natural and built environments and their interrelations. Higher-resolution input data leads to further spatial granularity in our understanding of disaster risk, enabling better targeting of disaster risk reduction policies and measures, and shedding light into previously unstudied spatial heterogeneities. Advancements pertaining to hazard, exposure, and vulnerability can be applied to inform a more nuanced, and thus better overall understanding of disaster risk.

Natural Hazard

A hazard can be defined as a "process, phenomenon or human activity that may cause loss of life, injury or other health impacts, property damage, social and economic disruption or environmental degradation." (UNGA, 2016). In this study, we focus only on natural hazards, which can be broadly categorised into hydrological, meteorological, climatological, and geophysical (Munich Re, 2018). Natural hazards may be either widespread or localised. The spatial footprint of an event is often impacted by the topology of the natural and built environments (e.g. elevation and slope, but also infrastructure influencing the hazard extent). High-resolution data generated by means of remote sensing is an essential input in models detailing hazards such as flood and landslide and is fundamental for understanding the spatial extent of both rapid-onset (e.g. landslide, flash flooding) and slow-onset (e.g. droughts, monsoon flooding) hazard events (Coppin et al., 2004; Aggarwal et al., 2016).

Conventional monitoring systems that collect in situ data of hydrological hazards (e.g. flow, precipitation, water level) are often limited in their applicability by insufficient spatial and temporal coverage (Van Dijk and Renzullo, 2011). Furthermore, the costs associated with collecting and maintaining such data can become a binding constraint for many countries (Klemas, 2015). Timely and large-scale satellite remote sensing has emerged as a viable alternative or supplement to in situ hydrological observations especially over ungauged regions (Khan et al., 2010). Satellite data can be utilised not only for flood detection, but also to improve flood modelling by incorporating high-resolution digital elevation models and

complementing hydrologic data assimilation (Revilla-Romero et al., 2015; Schumann, 2015; Tanaka et al., 2019).

Technologies that collect Synthetic Aperture Radar (SAR) data have increasingly proven effective for monitoring changes in surface water (Musa et al., 2015) as well as landslide hazard (Kang et al., 2019; Jung and Yun, 2020). Notably, these technologies are effective in all-weather conditions, both day and night, and they can also detect standing water beneath vegetation canopies, an important attribute for flood inundation over dense vegetated land (White et al., 2015; Martinis and Rieke, 2015; Martinis et al., 2018). SAR-based sensors record the backscatter characteristics and roughness of different surface features. Because open water bodies typically exhibit lower backscatter coefficients, stagnant water tends to appear darker in the recorded image compared to non-water areas which exhibit a higher backscatter (Brisco, 2015; Martinis and Rieke, 2015; Anusha and Bharathi, 2019; Cao et al., 2019).

Exposure

Exposure refers to "the situation of people, infrastructure, housing, production capacities and other tangible human assets located in hazard-prone areas" (UN General Assembly, 2016). Assets are often clustered in cities and near natural ports of connectivity (e.g. rivers and oceans). The ever-decreasing costs of high-resolution remotely sensed data, coupled with advances in storage capabilities and novel machine learning techniques, have driven a growth in our ability to understand how global exposure relates to population and assets. Released in 2005, the Gridded Population of the World, Version 3 (GPWv3) (CIESIN – Columbia University; FAO; CIAT, 2005) was one of the first datasets to provide global gridded population estimates at approximately 5-kilometer resolution. In the last decade, additional

datasets have been developed to characterise the spatial distribution of the global population, including WorldPop (Gaughan et al., 2013; Tatem, 2017) which provides estimates at 100-meter resolution and the High-Resolution Settlement Layer (HRSL) (Facebook Connectivity Lab and CIESIN – Colombia University, 2016) which provides estimates at approximately 30-meter resolution.

Methodologies often differ dramatically for estimating and modelling population on Earth (Leyk et al., 2019). For example, HRSL identifies potentially populated areas by applying a binary mask to proportionally reallocate populations from census aggregates (Tiecke et al., 2017). WorldPop, on the other hand, applies a weighted dasymetric approach that relies on a random forest model inclusive of many ancillary variables (Stevens et al., 2015). The common thread between existing global population estimation models is that they rely, at least to some extent, on remotely sensed derived data, which is further processed and analysed by means of different types of machine learning techniques. In addition to population exposure, it is also important to estimate the exposure of critical infrastructure and assets (e.g. building footprints, road networks, etc.) to different types of hazards. The Global Urban Footprint (GUF), which generates a settlement layer at approximately 12-meter resolution (75-meter resolution for public usage) using images from TerraSAR X and TanDEM X (Esch et al., 2011; Esch et al., 2017; Esch et al., 2018) is only one example to how remotely sensed data can be utilised for asset mapping.

Taking a step forward to look at the potential value of exposed assets, Gunasekera et al. (2015) develop a country level model at approximately 1-kilometer resolution using only openly available data that captures attributes relating to building inventory stock, non-building stock infrastructure density and value, as well as spatial and sectoral disaggregation of Gross Domestic Product (GDP). In addition to

remotely sensed observations, Volunteered Geographic Information (VGI) such as OpenStreetMap, provide a supplement to the known attributes of asset features (e.g. number of building stories). Through machine learning techniques, VGI can be paired with remotely sensed data to generate exposure estimates and support disaster risk management efforts (Johnson and Iizuka, 2016; Chen and Zipf, 2017; Goldblatt et al., 2020).

Vulnerability

Vulnerability refers to "the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards." (UN General Assembly, 2016). Vulnerabilities of populations are not spatially homogenous, with poorer populations noted to experience a "poverty exposure bias" for droughts and urban floods (Winsemius et al., 2015). Additionally, poorer populations may suffer disproportionately from disasters due to a greater share of their total assets being exposed to hazard events and more vulnerable overall building stock (Hallegatte et al., 2017). This was the case, for example, in the 2015 Nepal Earthquake, where most damage occurred in rural housing structures comprised of low-strength masonry (Lallemant et al., 2017).

Asset vulnerability is particularly challenging to monitor, because it often depends on small-scale factors pertaining to both the natural and built environments. Nonetheless, remotely sensed observations are continuing to provide new insights. Change detection techniques are used to refine models detailing liquefaction susceptibility for seismic risk which rely on spatially informed variables such as distance to water bodies, water table depth, and precipitation (Zhu et al., 2017). Vulnerability of structural assets to hazards can be estimated by means of high-

resolution lidar remote sensing that can characterise vulnerability based off building stock attributes (e.g. structure type) (Lu et al., 2014). Further granularity can be added by incorporating additional attributes from other sources, for example, building types from OpenStreetMap, or façade classifications identified through training of convolutional neural networks on street view images (Kang et al., 2019).

Understanding the vulnerability of populations remains a challenge due to inherent privacy concerns regarding the release of granular socioeconomic data (Arzberger et al., 2004). However, with the increased adoption of geo-enabled computer-assisted personal interview technologies, targeted household surveys enable a more spatially nuanced understanding of population vulnerability as it pertains to disaster events (Erman et al., 2019). Additionally, increases in computational power, development of new algorithms, and the growing willingness of governments to release census microdata have enabled advancements in microsimulation techniques, population synthesis, and agent-based modelling, which can serve as workarounds to privacy constraints. Population synthesis, i.e. the use of algorithms to generate new and anonymous microdata for an entire "synthetic population" (Rubin, 1993; Beckman et al., 1996), is increasingly applied to evaluate the socioeconomic impact of policies and programmes (Williamson et al., 2002; Barrett et al., 2011; Barthélemy and Toint, 2013).

5.1.3. Informing coastal resilience policy and interventions in Bangladesh

The Bangladesh Coastal Zone (BCZ) is an area demarcated by the *Coastal Zone Policy* (Ministry of Water Resources, 2005), that covers 19 districts, 47,201 square kilometres and 710 kilometres of coastline and is home to approximately 42 million

people (Bangladesh Bureau of Statistics, 2012). According to the *Coastal Zone Policy*, the BCZ lags behind other parts of the country in socioeconomic development, it struggles to cope with natural disasters and the gradual deterioration of the environment, yet it has the potential to significantly contribute to national development. Existing challenges in the BCZ are likely to be exacerbated by the effects of climate change and associated sea-level rise. 62 percent of coastal lands are less than three meters above sea level, while 83 percent are less than five meters above sea level (Bangladesh Water Development Board, 2012). The BCZ is highly vulnerable to tropical cyclones and subsequent storm surges, which are projected in climate models to increase in frequency and intensity in Bangladesh due to climate change (Huq et al., 2010; Uddin et al., 2019).

Advancements in spatial data collection (e.g. remote sensing, geo-enabled surveys, street view imagery), analysis (e.g. machine learning, computer vision), and modelling (e.g. agent-based modelling, population synthesis) have led to a rapid increase in the quality and quantity of data available to inform localised disaster risk in the BCZ. These advancements provide an opportunity to transform policy decision-making through a more effective application of spatially targeted interventions aimed at increasing resilience to disaster events. Resilience is defined as "the ability of a system, community or society exposed to hazards to resist, absorb, accommodate, adapt to, transform and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions through risk management" (UN General Assembly, 2016).

This paper demonstrates the opportunities and constraints of high-resolution data for informing decision-making in the BCZ. We look specifically at Cyclone Fani by drawing together a range of recently available open data sources to analyse the cyclone hazard, exposure and vulnerability. After first making landfall near Puri, India on May 3, 2019 as a Category 4-equivalent tropical cyclone, Cyclone Fani proceeded in a north-easterly direction, eventually making landfall in Bangladesh on May 4, by which time it had weakened into a deep depression. Although the most severe damage occurred in India, Bangladesh was heavily affected by tidal inundation paired with high wind speeds in the southern coastal districts, high wind speeds in the western districts, and substantial rainfall across most of the country.

5.2. Methodology

To demonstrate how high-resolution data can be applied to augment disaster-related decision-making, policies, and investment planning, we examine the impacts of Cyclone Fani on the population residing in the western part of the BCZ. We utilise Sentinel-1 synthetic aperture radar (SAR) data to delineate flooded areas, WorldPop gridded population as a basis for estimating potentially flooded households, and sample microdata paired with aggregate tables from the 2011 Bangladesh National Census to derive a granular spatial understanding of socioeconomic characteristics indicative of household vulnerability. Our approach demonstrates how advancements in data collection, analysis, and modelling and the integration of spatially nuanced risk information can be used to inform policy decision-making and investment planning. We utilise only open, free and publicly available data sources,

demonstrating that this type of analysis does not necessarily have to be a costly endeavour.

Our methodology entails the following three steps: (i) developing a granular estimation of flood delineation to enhance our understanding of Cyclone Fani's extent; (ii) spatially disaggregating a gridded population layer into a household point layer (unique x, y coordinates for each household) to enhance our understanding of the population exposed to Cyclone Fani; and, (iii) generating a population of synthetic households and joining their structural and socioeconomic attributes to the household point feature layer to provide insight into the relative vulnerabilities of exposed and non-exposed populations.

Although some level of damage was reported across 28 districts (Needs Assessment Working Group, 2019), our paper focuses on six contiguous districts in the western part of the BCZ where reports of flooding were most pervasive: (i) Bagerhat; (ii) Barguna; (iii) Khulna; (iv) Patuakhali; (v) Pirojpur; and, (vi) Satkhira (Fig. 22). Our study area spans nearly 16,000 square kilometres and contains approximately 9.5 million persons grouped into 2.2 million households (Bangladesh Bureau of Statistics, 2012).



Fig. 22 Study area comprised of (west to east) Satkhira, Khulna, Bagerhat, Pirojpur, Barguna, and Patuakhali districts

5.2.1. Developing a granular flood extent layer for Cyclone Fani

To gain a better understanding of the spatial extent of damages caused by Cyclone Fani, we utilise Sentinel-1 synthetic aperture radar (SAR) data to detect flooded areas.

We apply an image differencing technique (i.e. comparing images that were captured during and before the flood event) to delineate potential flooded areas in the six Fani-affected districts within our study area. The analysis is run in Google Earth Engine (GEE), a platform that leverages cloud-computing services to achieve planetary-scale utility. GEE has been previously used for flood detection and surface water mapping applications (Tang et al., 2016; Markert et al. 2018; Uddin et al. 2019).

Creating Sentinel-1 base and event images

We rely on Sentinel-1 archive (available in GEE) to identify scenes captured before and immediately after Cyclone Fani made landfall in Bangladesh on May 4, 2019. We use Ground Range Detected (GRD) scenes that were processed to generate a calibrated, ortho-corrected product inclusive of thermal noise removal, radiometric calibration and terrain correction. Each scene contains either one or two out of four possible polarisation bands: single band VV, single band HH, dual band VV+VH, or dual band HH+HV. In this study we rely on single co-polarisation, vertical transmit/vertical receive data (VV). VV polarisation has been successfully used for flood mapping in previous studies (Schlaffer et al., 2015), suggesting a slight advantage over other polarisation modes (Twele et al., 2016; Clement et al., 2018).

Base Image: We mosaic two Sentinel-1 scenes that were captured over the study area on April 28, 2019 (one week before landfall) to create a single Base Image BI:

- S1_GRD/S1A_IW_GRDH_1SDV_20190428T235604_20190428T235629_026997_030A0E_72BD
- S1_GRD/S1A_IW_GRDH_1SDV_20190428T235629_20190428T235654_026997_030A0E_FDF3

Event Image: We mosaic three Sentinel-1 scenes that were captured over the area of analysis on May 4, 2019 to create a single Event Image EI:

- S1_GRD/S1A_IW_GRDH_1SDV_20190504T000422_20190504T000447_027070_030CB5_863C
- S1_GRD/S1B_IW_GRDH_1SDV_20190504T235513_20190504T235542_016101_01E49E_F422
- S1_GRD/S1B_IW_GRDH_1SDV_20190504T235542_20190504T235616_016101_01E49E_4F22

Creating our difference image

Large portions of Sentinel-1 scenes over the study area cover permanent bodies of water (the Bay of Bengal, riverways, etc.) which we want to exclude from the analysis. We rely on the Joint Research Centre (JRC) Yearly Water Classification History dataset (version 1.1) (Pekel et al., 2016) (also available in GEE) to identify Permanent Water PW bodies, which we remove from EI and BI prior to analysis. We next calculate a Difference Image $DIFF$ by subtracting the value of each pixel in BI from the value of the overlapping pixel in EI:

$$EI_i - BI_i = DIFF_i$$

Where EI_i is the value of pixel *i* in the EI, BI_i is the value of pixel *i* in the BI, and $DIFF_i$ is the difference between the EI and the BI for pixel i.

To reduce the complexity of the analysis and the "noise" in the data, we smooth $DIFF$ by calculating a morphological reducer over the pixels in the image.

We reduce speckle noise by calculating a median value between neighbouring pixels (we use a smoothing kernel with a radius of 30 meters). The result of this step is a Smoothed Difference Image $SDIFF$ which contains fewer isolated pixels that could potentially be misclassified in the subsequent steps. Fig. 23 presents $DIFF$ and $SDIFF$ to highlight before and after smoothing.

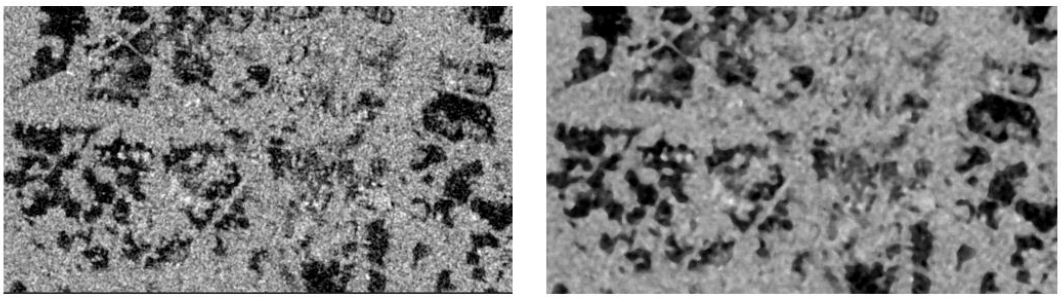


Fig. 23 Difference Image $DIFF$ (left) and Smoothed Difference Image $SDIFF$ (right)

Create a binary map of observed flooded areas

The final step in the analysis involves the conversion of $SDIFF$ into a binary map of observed flooded areas $FLOOD$ (i.e. flooded ("1") /not flooded ("0")). We develop a thresholding technique that relies on the relative values (a percentile value) of the pixels in the study area.

To detect potential flooding in both open spaces and in urban settings, we create two images: one image includes pixels that exhibit a decrease in the backscatter signal between the event and the base image DEC and one image which includes pixels that exhibit an increase in the backscatter signal between the event and the base image INC:

$$SDIFF = \begin{cases} DEC & SDIFF_i < 0 \\ INC & SDIFF_i > 0 \end{cases}$$

Where $SDIFF$ is the Smoothed Difference Image, DEC is an image which includes pixels i in the $SDIFF$ that exhibit a decrease in the backscatter value and INC is an image which includes pixels i in the $SDIFF$ that exhibit an increase in the backscatter value.

We convert the DEC and the INC images into a binary map of observed potentially flooded areas $FLOOD$. We do this by calculating a threshold, which is defined as a percentile of the value of all the pixels i in the study area (area of interest, aoi). A pixel i is marked as flooded if its value is either below the 1st percentile of all the pixels in the DEC image or above the 95th percentile of all pixels in the INC image. We set this threshold based on a visual assessment of the resulted flood map against the difference image:

$$FLOOD = \begin{cases} DEC_i < p(DEC(aoi), 1) \\ INC_i > p(INC(aoi), 95) \end{cases}$$

5.2.2. Spatially disaggregating household exposure

To gain a deeper understanding of the population exposed to flooding due to Cyclone Fani, we generate a point layer containing a unique location (x and y coordinates) representing the estimated location of each of the 2.2 million households located within the six districts in our study area. To do so, we rely on the Bangladesh WorldPop dataset for 2011 (www.worldpop.org) as a base population density layer. WorldPop provides estimated population density at approximately 100-meter resolution obtained using a Random Forest-based dasymetric mapping approach (Stevens et al., 2015). In Bangladesh, 21 spatial covariates are considered in the WorldPop estimates, including night-time lights, slope, elevation, and distance to

features of both the built environment (e.g. urban areas, major road intersections) and the natural environment (e.g. inland waterbodies, sparsely vegetated areas). We have chosen to use WorldPop over other gridded population layers due to its focus on country-specific data, including the usage of union level census data in its model.

Moving from gridded population counts to household point features

First, we convert the WorldPop dataset (approximately 100-meter resolution) from population counts (a float number) to household counts (an integer). For each union (admin level 4), the average household size is calculated from the 2011 Bangladesh National Census (Bangladesh Bureau of Statistics, 2012) and spatially joined to the WorldPop layer. The WorldPop float population counts are then divided by the average household size within the union, leaving float household counts. Next, the integer is separated from the decimal part. All integers are classified as households, while the decimal parts are used for weighted random allocation of remaining unallocated households within each union.

Each 100-meter grid cell now contains an integer attribute for the number of households located within the cell. To move from a polygon grid to a point layer, we create a dataset of randomly generated points within each cell corresponding to the integer value of households attributed to each cell (using the Create Random Points tool in ESRI ArcMap Version 10.6). Fig. 24 depicts approximately 58,000 predicted household point features across 550 square kilometres spanning Patharghata, Barguna Sadar, and Amtali *upazilas* in Barguna District. Clear clustering is visible around population centres of various area and density.

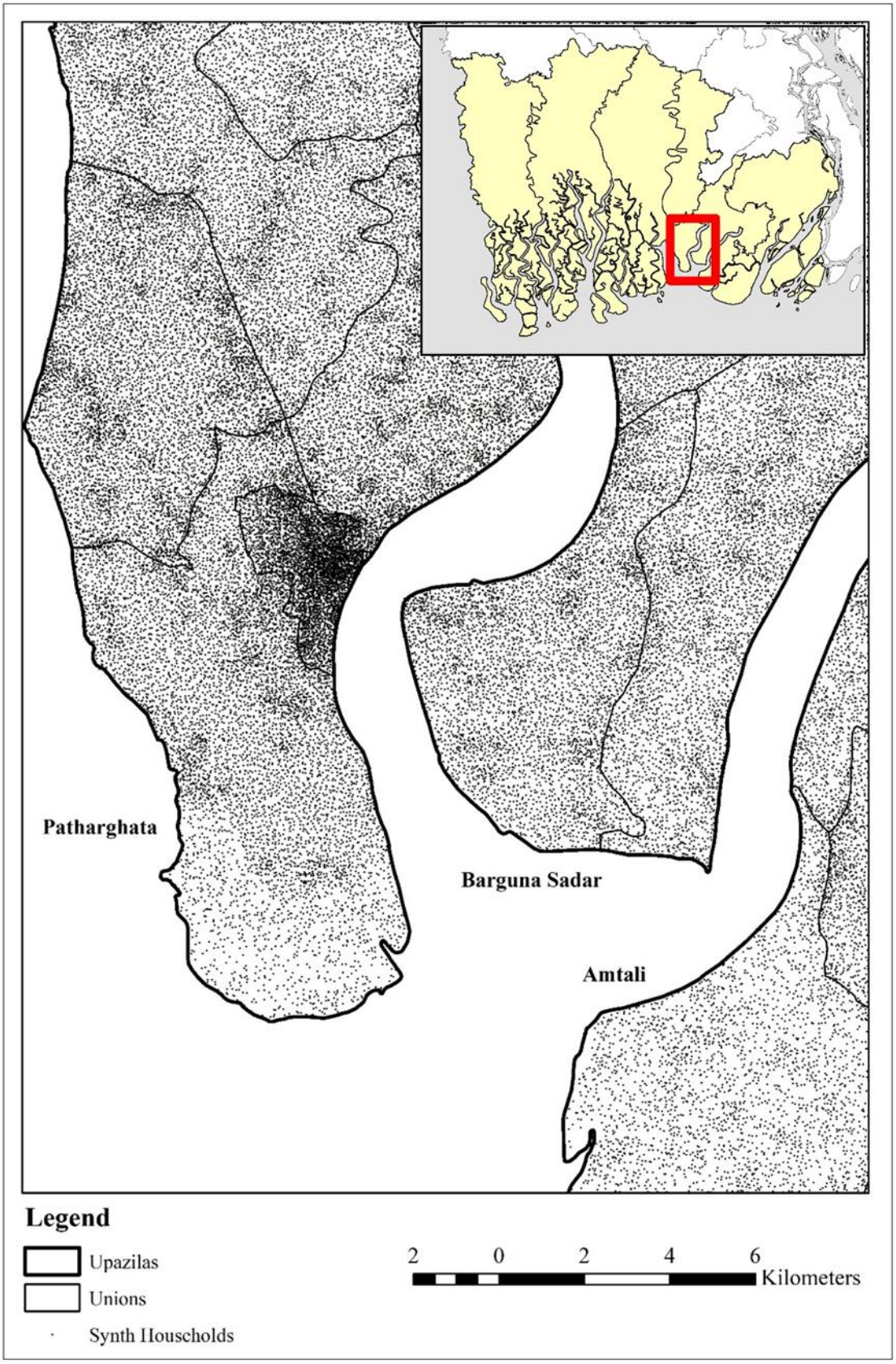


Fig. 24 Point feature class of predicted household locations spanning Patharghata, Barguna Sadar, and Amtali *upazilas* in Barguna District

5.2.3. Generating synthetic socioeconomic attributes as proxies for vulnerability

Generating high resolution estimates of population vulnerability over large areas has been one of the greatest challenges in spatial risk analysis. Synthetic socioeconomic attributes are increasingly used in disaster risk management to support dynamic agent-based simulation (Grinberger et al., 2017) and more generally to support bottom-up vulnerability assessments (Herman et al., 2016). To provide insight into the relative vulnerabilities of exposed and non-exposed populations in the study area, we generate synthetic structural and socioeconomic microdata for each of the 9.5 million persons, grouped into 2.2 million households within the six districts of our study area (utilising the framework provided by the R package SimPop (Templ et al., 2017)). Note that we do not attempt to construct an asset-based poverty metric as an overall proxy for welfare. Rather, we model certain variables which are indicative of different aspects of vulnerability. The synthetic population is first generated to correspond with population margins for each union, then each synthetic household is paired randomly (one-to-one relationship) with the derived household point features from the same union.

We rely on the following input data:

- International Public Use Microdata Series (IPUMS): 5% sample microdata from the 2011 Bangladesh National Census includes topology at the *upazila* level (admin level 3)
- Census Tables: Aggregate tables from the 2011 Bangladesh National Census at the union level (admin level 4)
- Demographic and Health Survey (DHS) Standard Data Set: 2011 DHS data set for Bangladesh

Correcting the census data set for age heaping and missing infants

The original census data shows clear age heaping as well as an underrepresentation of infants (defined as between the ages of 0 and 1). These errors are commonly reported in censuses and require correcting to be more representative of the true age distribution of the population (Stockwell and Wicks, 1974). The infants are imputed using variables which are present in both the census sample and the DHS data set. The imputed infants inherit all household variables from the household to which they were assigned and all personal variables by random draws from the set of infants already present in the census sample.

Generating a synthetic population at the upazila level

All steps of the synthetic data generation are split by the stratification variable *upazila*. We have chosen to stratify our data by *upazila* so that we can further stratify our data to the union level in subsequent steps. Thus, all models are estimated independently within each *upazila*.

Basic structural variables: To generate a realistic base for further modelling steps, the following basic variables for each household and household member are sampled from the Census microdata in simPop and kept fixed:

- Sex
- Age
- Relationship to head of household
- Urban/Rural
- Upazila

Household-level variables: All available variables which are defined on a household level are modelled and appended sequentially to basic structural variables:

- Household Type (general or institutional)
- House Structure Type (pucka, semi-pucka, kutcha, jhupri)
- Tenancy Status (owned, renting, free/usufruct)
- Water Source (tap, tube-well, other)
- Electricity (yes, no)
- Toilet Type (sanitary with water seal, sanitary without water seal, non-sanitary, none)

Individual-level variables: All available variables which are defined on an individual level are modelled and appended sequentially to basic structural variables and household-level variables: school

- Years of Schooling
- Literate (yes, no)
- Level of Education Completed (less than primary, primary, secondary, university)
- Employment Status (employed, unemployed, inactive)
- Industry (agriculture, industry, services)

Calibrating the synthetic population at the union level

Initially, generated households are randomly assigned to unions (admin 4) within each *upazila* (admin level 3), matching the total number of households listed per union in the census tables. In the next step, and in order to get a more realistic distribution of households with certain properties at the union level, calibration is performed through simulated annealing. Census aggregate data at the union level for household and

individual attributes are used to generate a more representative distribution of households at the union level within each *upazila*.

Household level:

- House Structure Type
- Water Source
- Toilet Type
- Tenancy Status
- Electricity

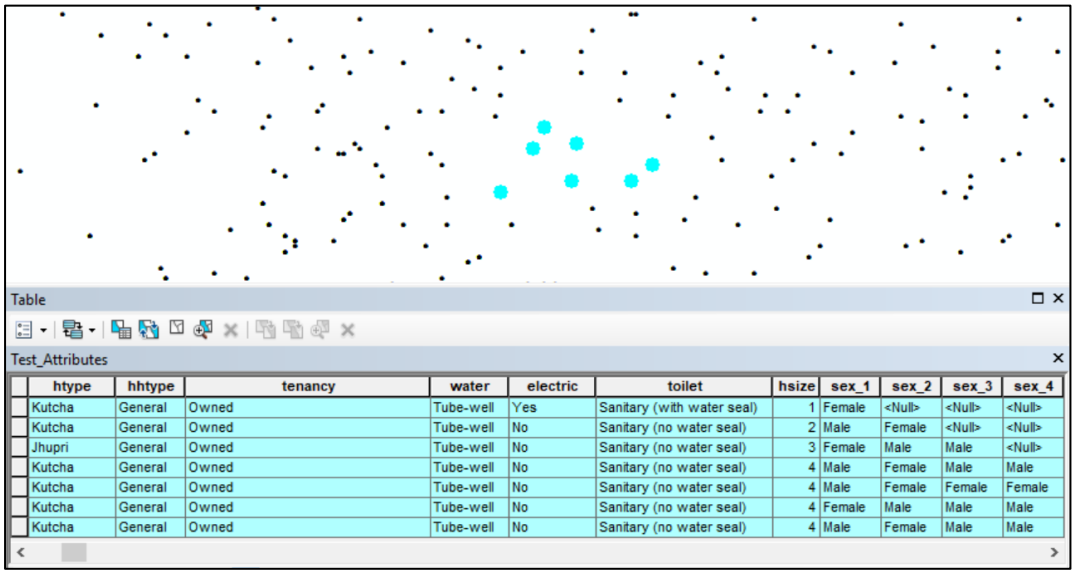
Individual level:

- Sex
- Age
- Literate

Joining synthetic structural and socioeconomic attributes to household points features

The households point layer is spatially joined with the administrative boundaries (from the 2011 Bangladesh National Census) to associate a unique ID which is used to randomly join the point feature class with discrete synthetic households within each union. Each household point feature now contains synthetic structural and socioeconomic attributes for each household as well as for all individual persons associated with that household. Fig. 25 details selected household point features and a sample of their associated synthetic attributes. Each row represents one synthetic household. Each column represents attributes associated with either the entire

household (e.g. htype, tenancy, etc.) or individual household members (e.g. sex_1, sex_2, sex_3, sex_4).



htype	hhtype	tenancy	water	electric	toilet	hsize	sex_1	sex_2	sex_3	sex_4
Kutcha	General	Owned	Tube-well	Yes	Sanitary (with water seal)	1	Female	<Null>	<Null>	<Null>
Kutcha	General	Owned	Tube-well	No	Sanitary (no water seal)	2	Male	Female	<Null>	<Null>
Jhupri	General	Owned	Tube-well	No	Sanitary (no water seal)	3	Female	Male	Male	<Null>
Kutcha	General	Owned	Tube-well	No	Sanitary (no water seal)	4	Male	Female	Male	Male
Kutcha	General	Owned	Tube-well	No	Sanitary (no water seal)	4	Male	Female	Female	Female
Kutcha	General	Owned	Tube-well	No	Sanitary (no water seal)	4	Female	Male	Male	Male
Kutcha	General	Owned	Tube-well	No	Sanitary (no water seal)	4	Male	Female	Male	Male

Fig. 25 Point feature class of synthetic households and associated attribute table of selected features in Barguna District

5.3. Results

Fig. 26 provides a visual snapshot of a small section of the study area – depicting approximately 172,000 synthetic households across 780 square kilometres of Bagerhat and Pirojpur Districts. Nearly 10,700 polygons depicting 46.3 square kilometres are observed as potentially flooded (highlighted in blue in Fig. 26).

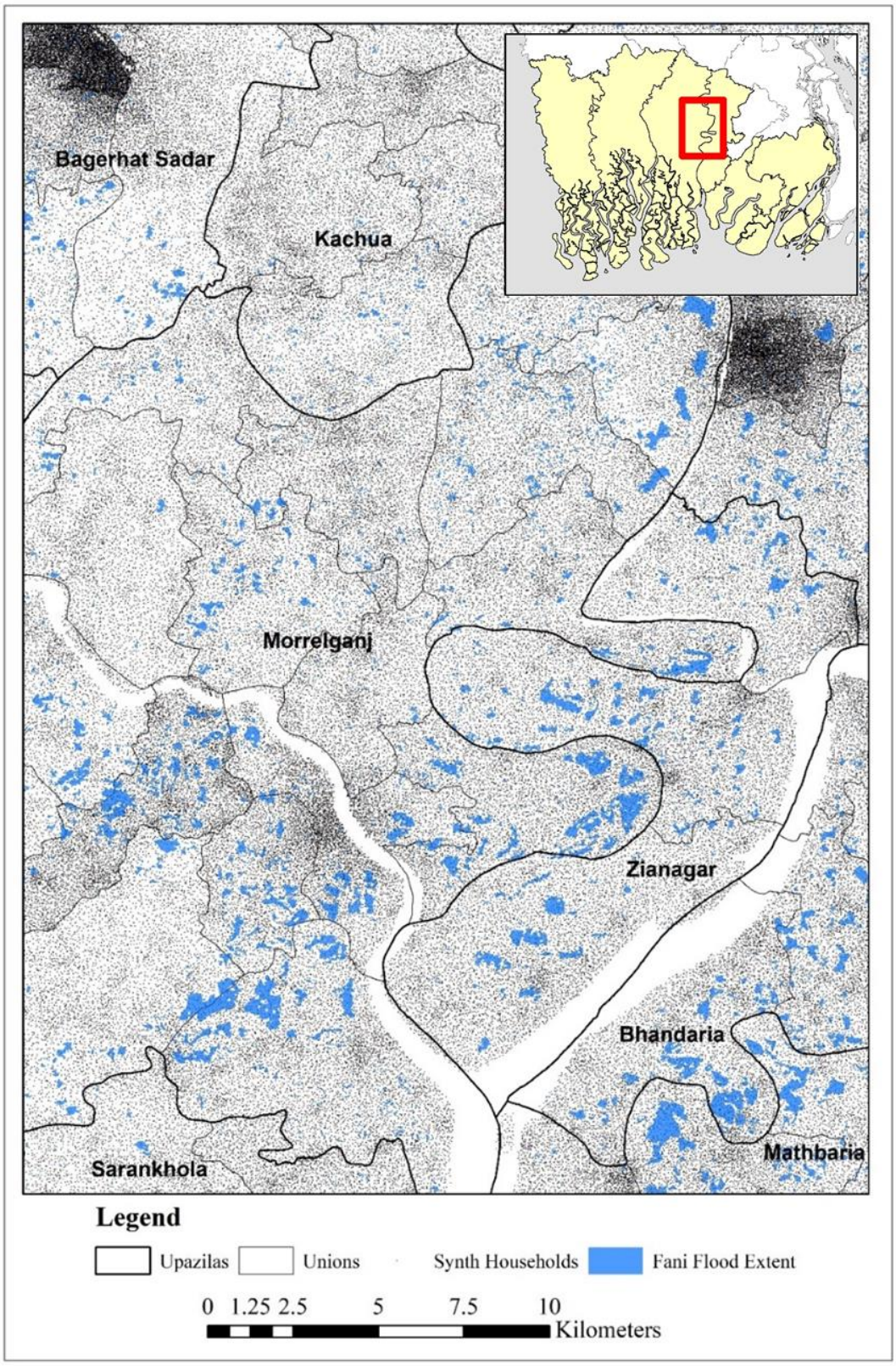


Fig. 26 Synthetic households (dark points) and Cyclone Fani flood extent from May 4, 2019 (in blue) spanning sections of Bagerhat and Pirojpur Districts

We compare the results with those reported in the *Cyclone Fani Joint Situation Analysis* released on May 13, 2019 by the Needs Assessment Working Group (NAWG), a coalition of NGOs, INGOs, and UN agencies and Clusters, under the direction and leadership of the Department of Disaster Management, Ministry of Disaster Management and Relief, Government of Bangladesh. We describe the results of this comparison in the next section.

5.3.1. Evaluation of observed hazard extent

Our results generated 334,837 polygons totalling 578 square kilometres of observed flooded areas (i.e. about 3.8 percent of the study area). We find a widespread partial inundation across the study area, with several more concentrated clusters. Only six out of the 49 populated *upazilas* show more than 10 percent of their area flooded (Table 14). At a more granular level, five of 436 unions show more than 20 percent of their area to be flooded (Table 15).

Table 14. *Upazilas with at least 10 percent observed flooded area*

Upazila	District	Flooded Area (sq. km)	Total Area (sq. km)	Percent Flooded
Chitalmari	Bagerhat	30.07	197.24	15.25
Satkhira Sadar	Satkhira	53.43	397.65	13.44
Kalaroa	Satkhira	26.43	232.77	11.35
Terokhada	Khulna	19.58	183.98	10.64
Kala Para	Patuakhali	45.97	444.16	10.35
Dighalia	Khulna	7.98	77.15	10.34

Table 15. *Unions with at least 20 percent observed flooded area*

Union	Upazila	District	Flooded Area (sq. km)	Total Area (sq. km)	Percent Flooded
Baikari	Satkhira Sadar	Satkhira	6.68	25.37	26.34
Hizla	Chitalmari	Bagerhat	6.47	25.29	25.60
Kuskhali	Satkhira Sadar	Satkhira	5.94	25.06	23.72
Bara Baria	Chitalmari	Bagerhat	7.43	35.25	21.07
Kalatala	Chitalmari	Bagerhat	5.63	27.08	20.80

The *Cyclone Fani Joint Situation Analysis* presents district-wise inundation estimates provided by the United Nations Operational Satellite Applications Programme (UNOSAT) under the United Nations Institute for Training and Research (UNITAR). Fig. 27 and Fig. 28 present visual comparisons between our detected (observed) flooded area and the flooded area provided by UNOSAT, while Fig. 29 compares the total flooded area per district for each of the two flood extent layers. Overall, our methodology results in more fragmented and smaller clusters of flooded areas compared to UNOSAT. This is particularly evident in the southern parts of Satkhira, Khulna, and Bagerhat over agriculture areas along the river delta as shown in Fig. 27. The differences are exemplified in Fig. 28, which presents a larger scale comparison of flooded agricultural areas in Khulna district. Our methodology does not detect many of the agriculture fields in this district as flooded and results in a more consistent distribution of flooded areas between districts (2.5-6.0% of the districts detected as flooded) compared to UNOSAT, which reports significant flooding primarily in the western districts of Bagerhat and Khulna (>9% of the area detected as flooded) and limited flooding in the eastern districts Barguna and Patuakhali (<1% of the area detected as flooded) (Fig. 29).

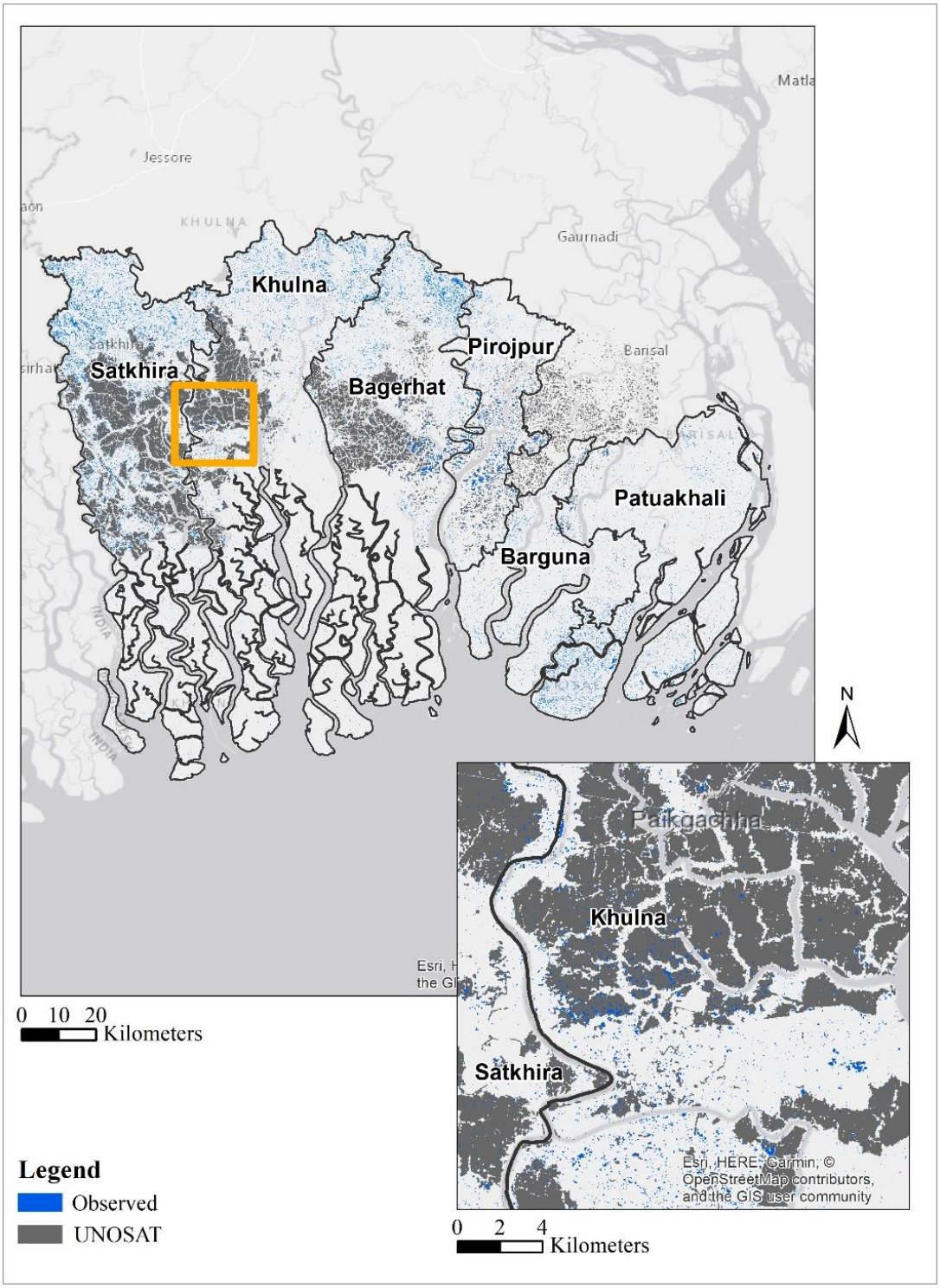


Fig. 27 Comparison of observed (blue) vs. UNOSAT (grey) flooded area

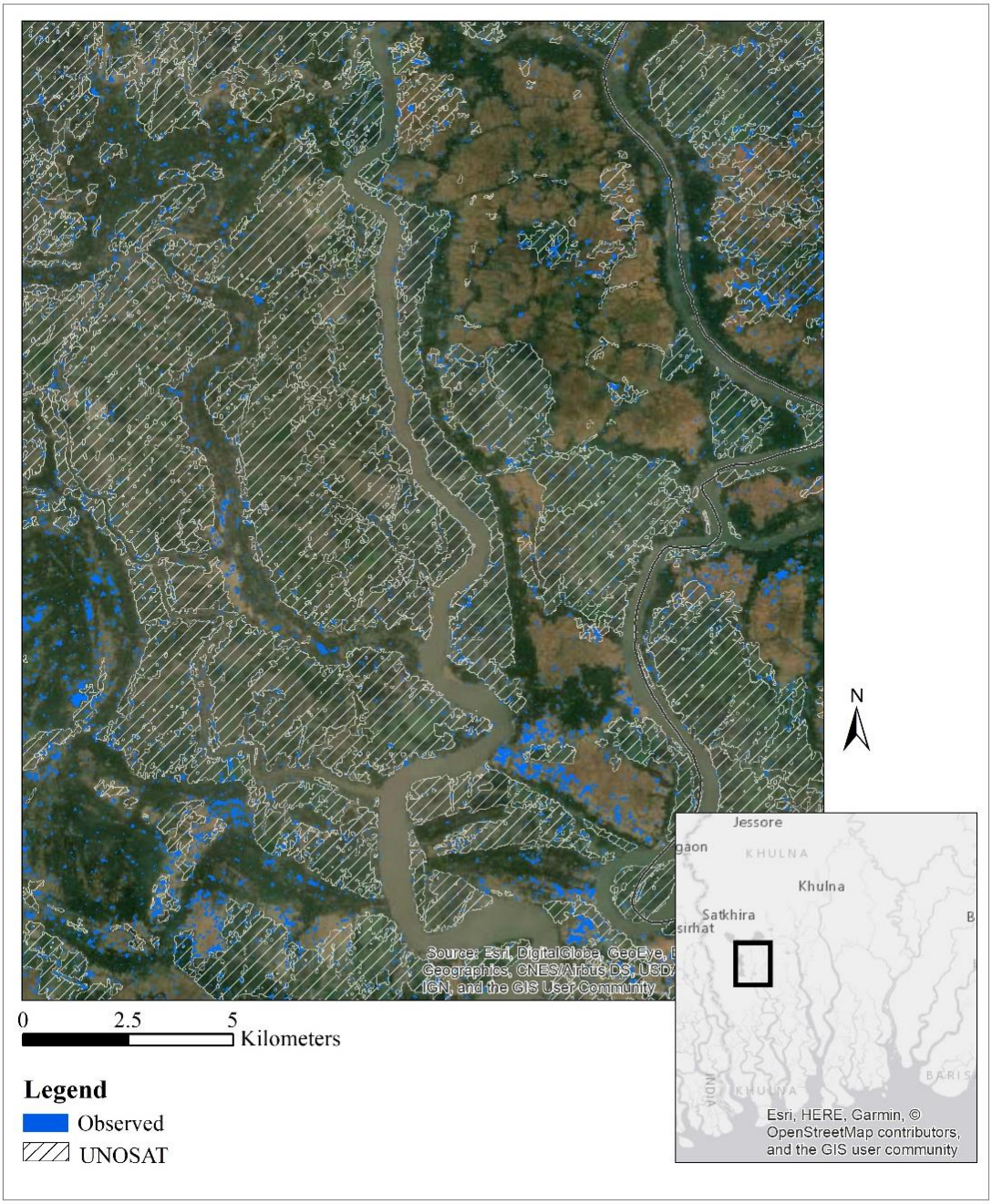


Fig. 28 Comparison of observed (blue) vs. UNOSAT (hatch pattern) flooded area

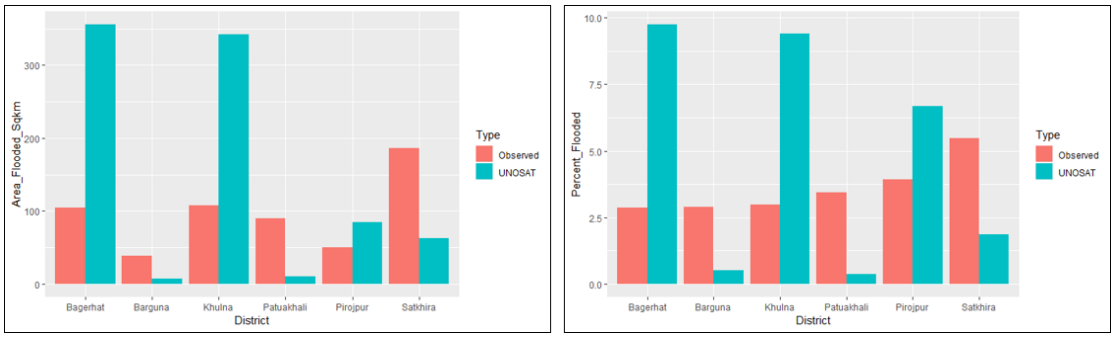


Fig. 29 District-wise comparison of observed vs. UNOSAT flooded area (left) and percentage of total district area (right)

The differences between our and UNOSAT predictions of flooding in Barguna and Patuakhali can be explained by different analytical and methodological factors. First, differences in the threshold set to declare a pixel as flooded may result in significant variations in the extent of detected flooding. As explained above, we calculate the threshold as a percentile value of all pixels in the study area (i.e. either 1^{st} or 95^{th} percentiles). Thus, our flood predictions are more "conservative", and a pixel is classified as flooded if its value is within the two extreme sides of the difference distribution tail. Second, differences in the smoothing and/or the generalisation algorithms may result in additional variations. The areas we detect as flooded are more spatially granular and less generalised compared to UNOSAT, which is characterised by larger clusters as clearly illustrated in Fig. 27 and Fig. 28. Third, some of the variations can be explained by the scenes used for the analysis. Flood events are dynamic by nature, and may present a different extent hourly, making the maximum flood extent extremely difficult to capture when scenes are captured only every few days. Our analysis mitigates this by creating a mosaic of three scenes covering the extent of our study area all captured within a span of five

minutes on May 4. UNOSAT use only one scene captured on May 4, which projects heavy flooding in western districts and limited flooding in eastern districts. Furthermore, the scene used by UNOSAT does not cover the spatial extent of coastal areas in Barguna and Patuakhali districts and thus does not consider tidal inundation. Our results are supported by the damages reported in the *Cyclone Fani Joint Situation Analysis* for Patuakhali district, which reported 45 percent of all damaged embankments, and Barguna district, which reported 44 percent of all damaged houses.

5.3.2. Evaluation of modelled populations exposed

Apart from flooded area, the *Cyclone Fani Joint Situation Analysis* is based on reports delivered by *upazilas,* districts and various line ministries in the days following the event, including a May 5 report from the National Disaster Response Coordination Cluster (NDRCC) detailing district-wise housing damages. Our modelled results are calculated through an overlay analysis, where household point features incorporated within our observed flooded area polygons are segregated and classified as "potentially flooded households".

Our modelled results show a more uniform distribution of potentially flooded households across districts (between 2.5% and 8.0% for all districts) compared to the NDRCC report, which shows a concentration of damaged households in Barguna District and none in nearby Bagerhat District (Fig. 30). Our estimates of flooded/impacted households do not consider damages caused by hazards other than direct flooding (e.g. wind damage), the prevalence of localised flood mitigation measures (e.g. elevated structures and berms), or concentrated mitigation failures (e.g. proximity to an embankment breach). They only represent an indicative

representation of where damages may have been incurred through a consistent general geospatial methodology.

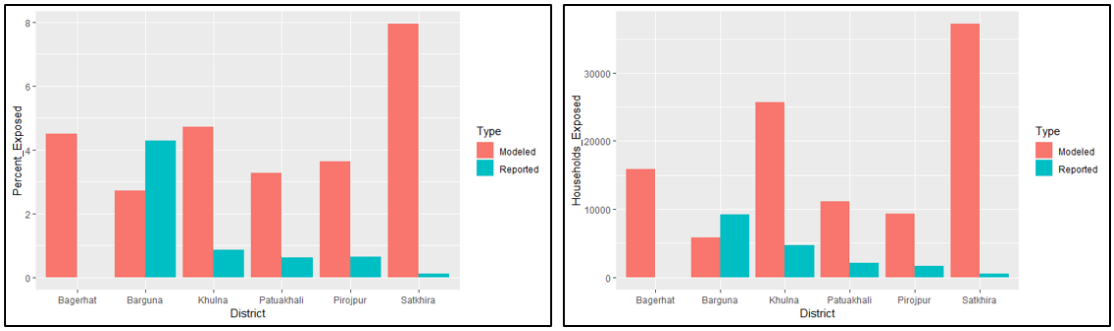


Fig. 30 District-wise comparison of modelled potentially flooded households vs. NDRCC reported damaged households (left) and percentage of total district population (right)

NDRCC reports represent an approach for data collection entirely reliant on locally reported estimates, which by its nature, is prone to subjectivity and methodological errors (e.g. difference in interpretation of "partially damaged") or to direct errors or lags in reporting (e.g. unlikely that Bagerhat District had no damaged households when reports for neighbouring districts were both in the thousands). An example of subjectivity is in the collection of "SOS forms" (which quickly estimate damages and needs) volunteered from *upazilas* following the event. The NAWG only received SOS forms from 35 *upazilas*, of which only 14 reported that they had been affected. Many of the stated numbers for damaged households were rounded to the nearest hundred indicating a further lack of precision in the process.

5.3.3. Evaluation of modelled vulnerability of populations exposed

To some extent, the *Cyclone Fani Joint Situation Analysis* presents limited information regarding the socioeconomic characteristics of affected households beyond a breakdown by sex. This is logical considering the noted limitations surrounding available data for populations exposed. However, our methodology allows us to take a closer look at the socioeconomic characteristics of potentially flooded households. Fig 31 provides a summary of selected socioeconomic characteristics from our modelled potentially flooded households data.

The proportional distribution of modelled attributes by district provides a proxy for understanding the relative vulnerability of exposed households. For example, certain structure designs may be better suited to withstand flooding than others (Islam et al., 2015). If *Jhupri* structures (jute and tree leaves) fare worse on average than *Pucka* structures (brick and concrete) and are primarily constructed by households in poorer rural areas, one can infer a relatively higher level of vulnerability for households with *Jhupri* structures.

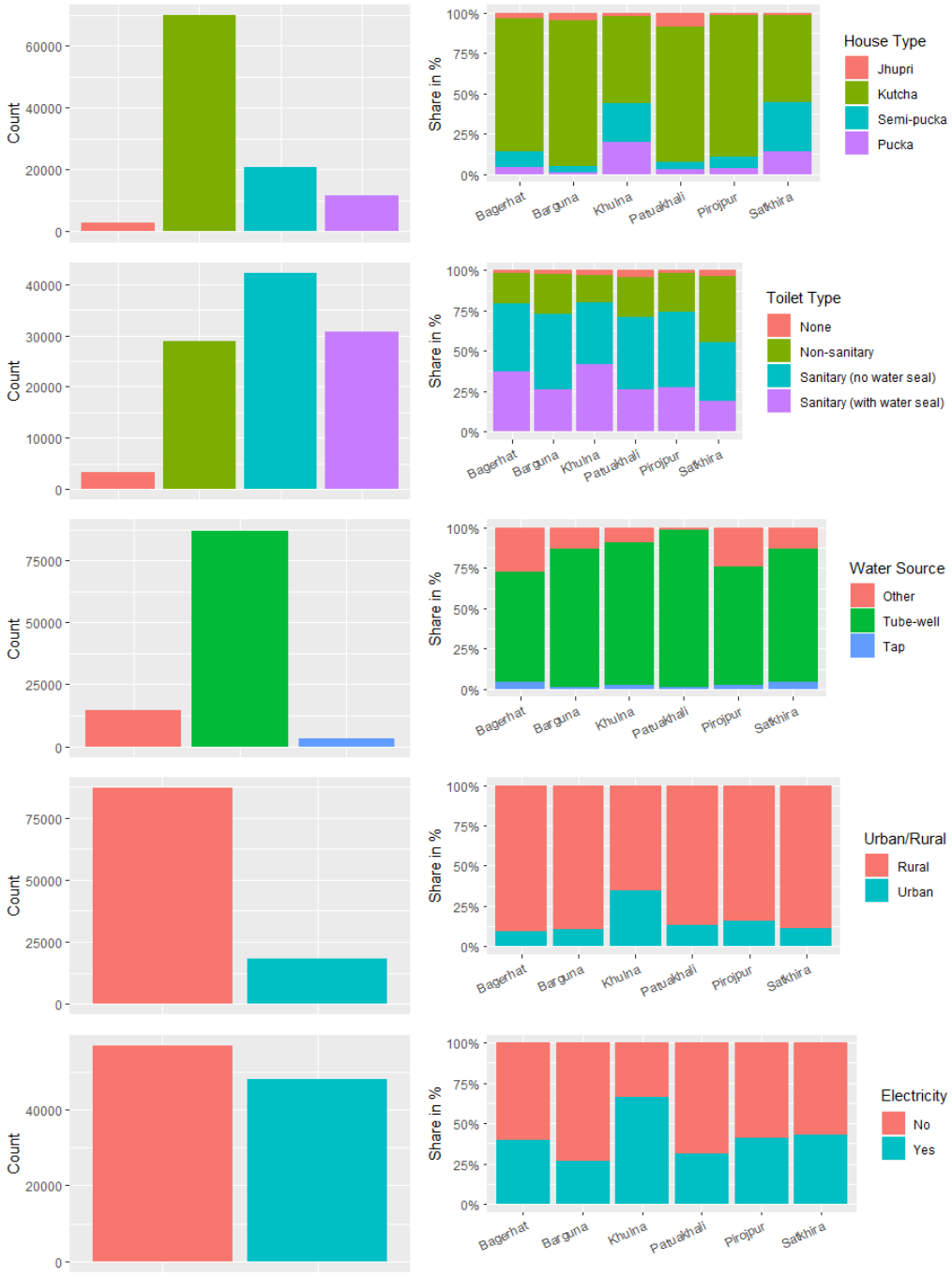


Fig. 31 Total modelled potentially flooded households by variable (left) and proportional distribution of modelled by district (right)

Fig. 32 presents the difference between the distribution of modelled flooded and modelled non-flooded households over each socioeconomic variable. The differences in distributions across districts are no larger than a couple percentage points for each variable. This indicates no discernible “exposure bias” of modelled

potentially flooded households when compared to modelled non-flooded households, a conclusion consistent with that of Winsemius et al. (2015) for primarily rural flood events.

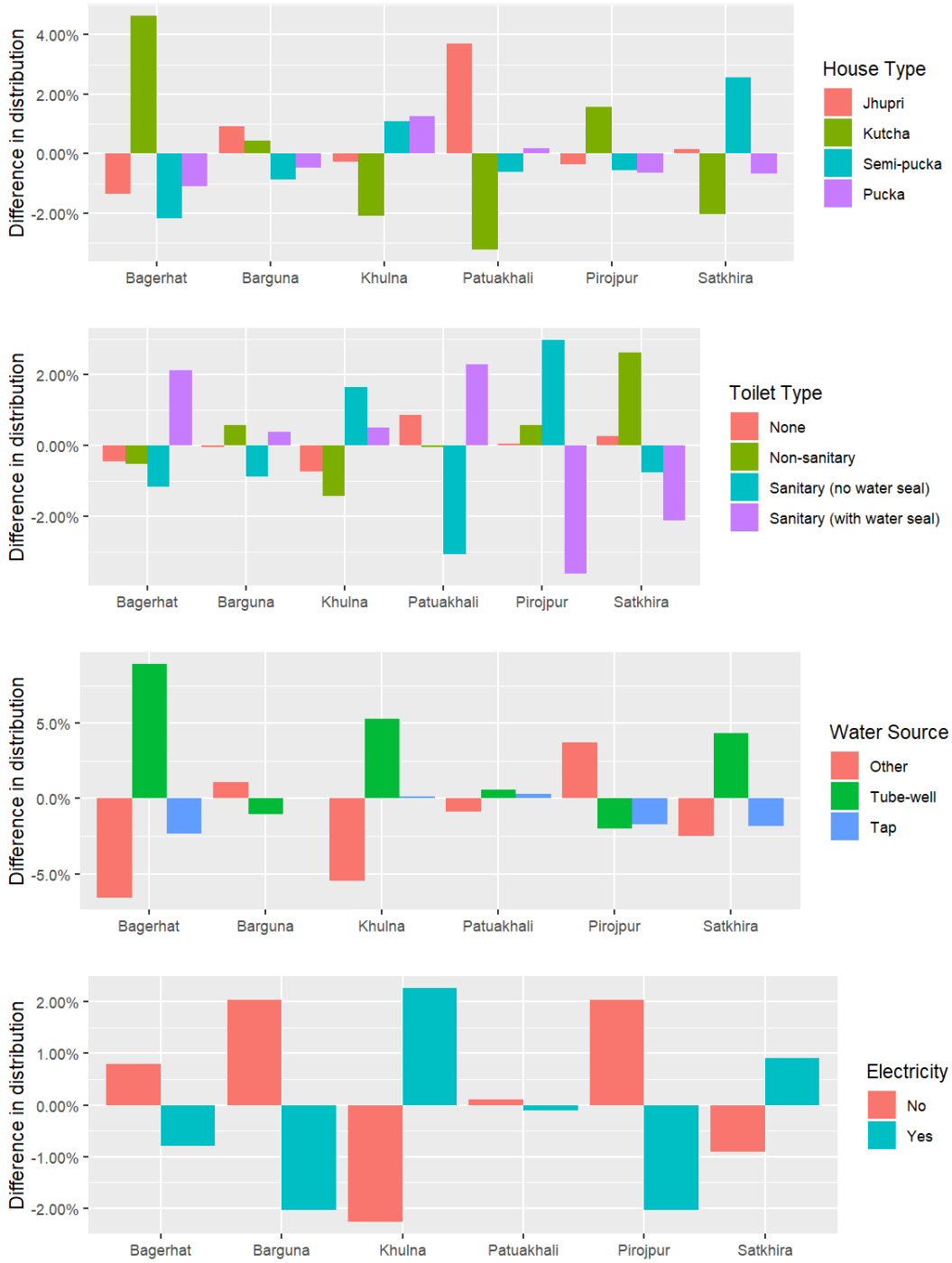


Fig. 32 Difference from the expected distribution for variables of modelled potentially flooded households

5.4. Discussion

5.4.1. Applications for augmenting disaster response and recovery efforts

Effective disaster response and recovery is contingent on accurate and timely data detailing affected areas *ex-post* (after the event). Until recently, efforts have focused on collection of in situ data and direct reports from those situated on the ground during a flood event. The *Standing Orders on Disaster* (Disaster Management Bureau, 2010) enshrined into law following the passage of the 2012 Disaster Management Act provides the legal mandate, management structure, and responsibilities for disaster response coordination groups at various levels of local government, including for each district, *upazila*, and union. The *Standings Orders on Disaster* represent a cascading system that relies on rapid collection of on the ground data and reports, strong participation from many local stakeholders, and strong coordination mechanisms to ensure that information can effectively flow upwards to higher levels of government. While the *Standing Orders on Disaster* presents the foundation for strong disaster response and recovery, our study highlights the potential to supplement the task of data collection through means of remote sensing and advanced modelling techniques.

Available through the European Space Agency's Copernicus Programme since 2016, Sentinel-1 provides SAR measurements which can be rapidly collected over large affected areas at day and night and across all weather conditions, without the need to rely on on-the-ground large scale data collection and coordination efforts. Relying on remotely sensed SAR measurements analysed through novel analytical techniques can eliminate much of the inherent subjectivity of ground reporting and enhance precision. Although SAR data has already been shown to be effective for

guiding response efforts for long duration monsoon flooding in Bangladesh, infrequent overpasses of satellites that carry SAR sensors has until recently limited its utility for shorter duration events (Duti et al., 2017). Recent advancements in remote sensing technologies and the increasing availability of satellites with SAR sensors, including non-proprietary data, provide new opportunities for exploitation of SAR data to understand cyclone-induced flood extents and to provide more frequent, granular updates of large-scale flood extent during major disasters.

Our study takes the transferrable data-based high-resolution spatial approach a step further by demonstrating a novel methodology for deriving synthetic households with unique geographic locations and socioeconomic attributes. Household exposure and vulnerability are modelled *ex-ante* (before the event), meaning that a simple overlay with the flood extent can provide fundamental insights into potential damages as well as the ability to target response efforts towards those populations which may be disproportionately affected. Our approach can be applied to better inform the response and recovery phases of a disaster risk management cycle, particularly at a central level (e.g. with the Ministry of Disaster Management and Relief) where quick access to reliable aggregate figures are essential for coordinating relief efforts and large-scale resource mobilisation. Modelled socioeconomic characteristics may provide utility in prioritising specific and targeted actions when coordinating relief efforts (e.g. decisions as to where to provide temporary shelter, water, or sanitation services).

Our study does have several limitations. Notably, there are no authoritative data sets detailing the extent of flooded areas beyond the volunteered reports submitted at the *upazila* level. This study exemplifies how slight changes in methodology can drastically change the flood extent of an event. Furthermore, the

synthetic population developed for this study reflects demographics from the 2011 Bangladesh National Census. Although one could project the data forward, we have decided against it for the purposes of this study as the intent was to produce rough estimates that could be used to quickly inform response and recovery rather than a detailed accounting of damages. Improving the effectiveness of response and recovery efforts through rapid insights and high-level targeting of efforts can enable affected populations to bounce back quicker from a disaster event, enhancing the overall resilience of communities located within the BCZ.

5.4.2. Applications for prioritising risk-mitigating investments and enhancing preparedness

In addition to providing utility in the response and recovery phases of the disaster risk management cycle, our approach can be applied to inform *ex-ante* investments and associated policies focused on enhancing the resilience of the BCZ across the mitigation and preparedness phases. Because our flood delineation layer represents a single event, the data must be integrated into a broader hazard model (e.g. through data assimilation and data fusion) to be of broader use for informing *ex-ante* decision-making. Conversely, our population exposure and population vulnerability data are generated across the study area and can be integrated directly into existing flood risk models. A better baseline understanding of where populations live, and their associated socioeconomic characteristics, can be applied to inform decision-making around large-scale risk-mitigating capital investments in the BCZ.

The BCZ contains 139 polders (49 sea facing) and 2,900 drainage control structures that aim to protect approximately 8 million people (out of 35 million people across the entire BCZ) and 12,000 kilometres of predominantly agricultural land (Kibria and Khan, 2016). These polders were initially designed for the purpose of protecting land from tidal fluctuations and did not account for long-term risks and challenges from cyclonic storm surge events or subsidence, sedimentation, reduced river flows and other natural and human-induced geomorphological change. They were designed to maintain and make necessary improvements to the existing polders requires recurrent large-scale investments. Phase-1 of the Coastal Embankment Improvement Project (CEIP), invests in structural improvements of 17 polders with an estimated cost of $400 million, making it one of the Government of Bangladesh's largest projects (World Bank, 2013). The initial multicriteria analysis used for selecting project polders considered physical factors such as the length that had been breached and eroded, hectares of drainage congestion, whether the polder was in a high-risk zone, and the opinion of stakeholders. When combined with spatially nuanced exposure models and risk-models that incorporate factors such as polder-breaching (Islam et al., 2019) the approach we propose in this study can enable future prioritisation exercises targeting polder improvements to further integrate population characteristics in addition to physical characteristics.

The same type of prioritisation could be applied to cyclone shelters, where the Government of Bangladesh is currently investing $375 million in the Multipurpose Disaster Shelter Project (MDSP). The MDSP invests in rehabilitating 450 existing shelters and constructing 550 new shelters (World Bank, 2014). However, an analysis conducted by the Local Government Engineering Department (LGED) concluded that a total of 7,124 shelters must be constructed by 2025 to improve the resilience of

communities across the BCZ. Our exposure data provides the baseline to perform an accessibility analysis of households to nearest existing shelters, as well as proposed locations for new shelters. Associated socioeconomic characteristics can serve to ensure that the most vulnerable populations are prioritised (e.g. those areas with high concentrations of *Jhupri* houses) when deciding the placement of new shelters or the upgrading of existing shelters.

To enhance preparedness, our approach and derived data can be used to better understand where populations may be most vulnerable. The modelled socioeconomic variables could contribute towards a multidimensional understanding of household poverty (Alkire et al., 2015) and socioeconomic resilience (Hallegatte et al., 2017). Taking this a step further would be to explore investments in disaster-linked social protection schemes, where vulnerable populations receive cash payments directly following a disaster to enable quicker recovery, thus enhancing their resilience in the occurrence of a disaster event. Additionally, preparedness-focused programmes such as the Cyclone Preparedness Programme (CPP) of the Bangladesh Red Crescent Society (BDRCS) can use our baseline information to inform early-warning systems, better target vulnerable communities, and ensure that they have the tools and capacity required to avoid damages and losses when an event occurs (Habib et al., 2016).

5.5. Conclusions

In this study we presented an integrated analytical approach for risk analysis in the case study of six districts in the BCZ that were impacted by Cyclone Fani. Granular estimations of hazard, exposure, and vulnerability are fundamental across the disaster risk management cycle and for developing informed policy and investments for

enhancing resilience across the BCZ. Our analysis demonstrates how combining a high-resolution estimate of the Cyclone Fani flood extent, a granular understanding of the population exposed, and modelled socioeconomic characteristics of those populations exposed could provide essential hazard and risk exposure estimates. Our analysis augmented the findings detailed in the *Cyclone Fani Joint Situation Analysis* in the following ways: (i) our observed hazard extent layer employed a more conservative pixel classification methodology to account for agricultural areas, resulting in more fragmented and smaller clusters of flooded area polygons and a more consistent distribution between districts when compared to estimates provided by UNOSAT; (ii) our modelled populations exposed layer provided an estimate of potentially flooded households which could be applied immediately to supplement locally reported estimates of damaged households; (iii) our modelled vulnerability of populations exposed provided novel insight into the structural and socioeconomic characteristics of households, enabling a basic understanding of the relative vulnerability of potentially flooded households across districts.

More broadly, the case study we presented exemplifies how developments in geospatial analysis can be applied to better inform disaster risk. We utilised three primary high-resolution and granular data inputs: (i) SAR data from Sentinel-1; (ii) gridded population estimates from WorldPop; and, (iii) structural and socioeconomic microdata and census tables acquired from the 2011 Bangladesh National Census. Sentinel-1 SAR sensors have collected data only since 2016 and offer the first consistently open high-resolution SAR data (other forms of targeted high-resolution proprietary SAR data, such as TerraSAR-X and RadarSAT-2, have been made openly available in the past following disaster events through the International Charter Space & Natural Disasters). WorldPop continues to improve and update its open population

estimates by refining its model and incorporating new spatial covariate layers as they become available. Lastly, the increasing availability of high-resolution sample microdata and advancements in population synthesis techniques enables new approaches to downscale population attributes, allowing for a more granular understanding of structural and socioeconomic characteristics. By applying these three data sources towards providing a better understanding of Cyclone Fani's impact on the BCZ, our analysis demonstrates how novel modelling approaches utilising high-resolution and granular geospatial data can provide added value across the disaster risk management cycle to inform policy decision-making and investment planning.

6. Concluding Remarks

6.1. Thesis summary

The primary aim of this thesis has been to evaluate the validity and effectiveness of emerging techniques for high-resolution geospatial analysis applied to disaster risk management. This is important for advancing our understanding of novel forms of high-resolution spatial insights and translating those insights into actions that can help developing countries overcome challenges pertaining the first priority of the *Sendai Framework for Disaster Risk Reduction 2015-2030 (SFDRR)*: understanding disaster risk.

Chapter 2 presented key definitions and concepts relevant to disaster risk management, outlined the evolution of disaster risk management as a global priority, and highlighted the critical importance of first understanding disaster risk before

making decisions across the disaster risk management cycle. The chapter also reviewed key technological and methodological advancements in areas of satellite Earth observation, localised data collection, and modelling techniques that have progressed our understanding of disaster risk. Three key breakthroughs relevant to the subsequent chapters were reviewed: (i) mapping population exposure; (ii) demonstrating more granular insights into population vulnerability; and, (iii) measuring the extent of flood hazard events.

Chapters 3, 4, and 5 each presented a peer-reviewed journal article that contributes towards achieving the aim of this thesis by addressing one of three research questions:

- Chapter 3: How accurate are emerging techniques for high-resolution global population modelling that are reliant on precise mapping of the built environment?
- Chapter 4: How can more granular and spatially nuanced insights be provided to inform population vulnerability in data scarce environments?
- Chapter 5: How can satellite remote sensing techniques be applied to better understand the impacts of natural hazards and supplement disaster risk management practices in developing countries?

Chapter 3 investigated the accuracy of recent advancements in global population modelling focused on precise mapping of the built environment. The accuracy of a range of methods were assessed that progressively increased the strength of binary masks derived from the built environment (land areas, all building footprints, residential building footprints) and of density variables relating to the

binary masking variable (building footprint areas, building volumes). The assessment was applied in New York City, where large spatial heterogeneities exist across confined geographic areas. Results confirmed that density variables directly relating to the binary masking variable are effective. However, their application requires careful consideration due to its propensity to amplify both positive results and errors. In order to mitigate these risks, they should be used in combination with additional density-informing ancillary variables.

Chapter 4 addressed the challenge posed by limited granularity and spatially nuanced insights in population vulnerability by proposing a methodology for generating spatially referenced synthetic households. The methodology results in an output vector layer of households represented as points with unique structural and socioeconomic attributes which can be applied towards a better understanding of population dynamics, particularly those characteristics which are geographically influenced and may not conform to administrative boundaries linked to aggregated census statistics. To demonstrate the effectiveness of the methodology, Chapter 4 provided a case study in Dumuria Upazila, Bangladesh, where an independent household survey was used to validate the location of the synthetic households and associated socioeconomic attributes. The study found that 98 percent of all surveyed households were within 100 metres of at least one synthetic household and the median distance observed was only 18 metres. The study also found that the age distribution of synthetic populations was within five percent of the surveyed population for each of the five age bands considered.

Lastly, Chapter 5 demonstrated how satellite remote sensing techniques to understand natural hazards can be integrated with the methodologies from previous chapters to target disaster response and recovery, as well as broader efforts to enhance

resilience. Chapter 5 demonstrated the usefulness of these methodologies in a cyclone risk analysis in an area of the Bangladesh Coastal Zone that is inhabited by 9.5 million people, using the 2019 Cyclone Fani as an example. The analysis offered new insights into the impact of the cyclone on populations by generating a high-resolution estimate of the flood extent, modelled point features of potentially flooded households, and modelled synthetic structural and socioeconomic characteristics of each household as a proxy for vulnerability. Chapter 5 demonstrated how high-resolution geospatial analysis could serve to supplement locally reported estimates of damaged households by providing a more consistent distribution of flood extent between districts and providing estimates of the relative vulnerability of potentially flooded households across districts.

6.2. Primary research contributions

The primary research contributions of this thesis concern the proposal and validation of methodologies reliant on recent advancements in high-resolution geospatial analysis, and their applicability towards improving disaster risk management in developing countries. Chapters 3, 4, and 5 each focused on one specific breakthrough in understanding disaster risk centred around aspects of population exposure (Chapter 3), population vulnerability (Chapter 4), and natural hazard events (Chapter 5). The primary research contributions within each chapter are summarised in the subsequent paragraphs.

Chapter 3 presents a simplified equation for global dasymetric population modelling that separates masking variables and density variables. Rather than proposing a novel methodology, the equation simplifies those proposed elsewhere (e.g. Mennis, 2003; Mennis and Hultgren, 2006; Sridharan and Qiu, 2013; Nagle et

al., 2014; Rebelo et al., 2019) to focus on two distinct objectives: (i) increasing the precision and accuracy of a strong binary mask to limit the spatial extent of settled populations; and, (ii) improving modelling techniques of ancillary variables to account for variations in density within the identified spatial extent. Chapter 3 also presents a critical assessment of population modelling techniques that heavily rely on derived characteristics of the built environment as a masking variable. A key example of this is the High-Resolution Settlement Layer (HRSL), which offers country-extent gridded population data at 30-metre resolution using building footprints delineated using computer vision techniques from 50-centimetre resolution satellite imagery (Tiecke et al., 2017). The assessment confirms that density variables directly relating to the binary masking variable are effective; however, their application requires careful consideration due to its propensity to amplify both positive results and errors.

Chapter 4 presents a novel methodology for generating spatially referenced synthetic households by combining population synthesis techniques for small area estimation (Templ et al., 2017) with dasymetric population modelling (Stevens et al., 2015). The methodology results in an output vector layer of households represented as points with unique structural and socioeconomic attributes which can be applied towards a better understanding of population dynamics, particularly those characteristics which are geographically influenced and may not conform to administrative boundaries linked to aggregated census statistics. The case study of Dumuria Upazila, Bangladesh, validates the methodology and provides an example of how georeferenced surveys can be applied towards error assessments in data-scarce regions, which remains a challenge in global population modelling (Wardrop et al., 2018; Leyk et al., 2019).

Chapter 5 presents a novel methodology for applying freely available synthetic aperture radar (SAR) data towards flood detection of large geographic areas immediately following a cyclone hazard event. The methodology resulted in enhanced precision, demonstrated by 334,827 unique polygons totalling 578 square kilometres of flooded areas in coastal districts of Bangladesh heavily affected by 2019 Cyclone Fani. The methodology contributes towards the growing body of literature pertaining to flood detection from Sentinel-1 SAR data (e.g. Martinis, 2017; Bioresita et al., 2018; Clement et al., 2018; Tsyganskaya et al., 2018; Uddin et al., 2019). The case study presented in Chapter 5 demonstrates how high-resolution geospatial analysis can supplement locally reported estimates of damages by providing quicker estimates and a means for identifying potential bias from self-reporting. Furthermore, insights into the socioeconomic characteristics of potentially affected households provide a layer of information not covered by the *Cyclone Fani Joint Situation Analysis* (Needs Assessment Working Group, 2019).

6.3. Practical Applications

In addition to research contributions centred around advancing global understanding of disaster risk, this thesis is structured to also consider practical applications of those advancements. This thesis contributes to the first priority for action of the SFDRR, understanding disaster risk, by addressing several of the key activities proposed to guide stakeholders at national and local levels, notably activities 1-6:

- Activity 1: “To promote the collection, analysis, management and use of relevant data and practical information and ensure its dissemination, taking into account the needs of different categories of users, as appropriate”

- Activity 2: “To encourage the use of and strengthening of baselines and periodically assess disaster risks, vulnerability, capacity, exposure, hazard characteristics and their possible sequential effects at the relevant social and spatial scale on ecosystems, in line with national circumstances”
- Activity 3: “To develop, periodically update and disseminate, as appropriate, location-based disaster risk information, including risk maps, to decision makers, the general public and communities at risk of exposure to disaster in an appropriate format by using, as applicable, geospatial information technology”
- Activity 4: “To systematically evaluate, record, share and publicly account for disaster losses and understand the economic, social, health, education, environmental and cultural heritage impacts, as appropriate, in the context of event-specific hazard-exposure and vulnerability information”
- Activity 5: “To make non-sensitive hazard-exposure, vulnerability, risk, disaster and loss-disaggregated information freely available and accessible, as appropriate”
- Activity 6: “To promote real time access to reliable data, make use of space and in situ information, including geographic information systems (GIS), and use information and communications technology innovations to enhance measurement tools and the collection, analysis and dissemination of data”

Chapters 3 and 4 focus on providing a better baseline understanding of population exposure and vulnerability. A better baseline understanding of where populations live, and their associated socioeconomic characteristics, can be applied to inform decision-making around large-scale risk-mitigating capital investments and enhancing preparedness. The case study presented in Chapter 5 highlights the utility

of this information for investment planning and prioritisation of coastal embankments and cyclone shelters across the Bangladesh Coastal Zone. The case study also highlights the potential applications of enhanced baseline population data towards enhancing preparedness through improved targeting of potential vulnerability hotspots for localised risk reduction efforts and early-warning systems.

The case study presented in Chapter 5 also integrates a hazard event, 2019 Cyclone Fani, and discusses the practical applications of quicker and more precise flood detection paired with an enhanced baseline understanding of population characteristics towards informing decision-making in the response and recovery phases of the disaster risk management cycle. Until recently, efforts in many developing countries have focused on collection of in situ data and direct reports from those situated on the ground during a flood event. Bangladesh presents a cascading system that relies on rapid collection of on the ground data and reports, participation from many local stakeholders, and strong coordination mechanisms to ensure that information can effectively flow upwards to higher levels of government. While in situ data collection may still present the foundation for strong disaster response and recovery efforts, Chapter 5 highlights the potential to supplement the task of data collection through means of remote sensing and advanced geospatial analysis.

6.4. Future Research

At 100-, 30-, or even 10-metre resolution, the resolution of global- or country-extent gridded data depicting settled (also referred to as “night-time”) populations can only be enhanced so much further. However, much more can be done in the future to incorporate advancements in satellite Earth observation, localised data collection, and

advanced modelling techniques to improve the accuracy of population estimates both in urban areas with high levels of spatial heterogeneity and rural areas with limited input data. This includes enhancing the precision and accuracy of masking variables to further limit the spatial extent of settled populations (e.g. incorporating building typology), providing additional density-informing attributes directly related to masking variables (e.g. building volumes), and including additional spatial covariates with more complex relationships. Furthermore, bottom-up approaches such as those that utilise targeted surveys (Weber et al., 2018) or mobile phone data (Deville et al., 2014) provide an opportunity to supplement population estimates in regions where baseline aggregates may not be up to date or fully trusted.

This thesis takes population modelling a step further by introducing a methodology for generating spatially referenced synthetic households as a vector layer of point features with associated socioeconomic and structural attributes for each individual and household. More research is clearly required to understand population vulnerability at scale in developing countries. This was exemplified by the lack of adequate data pertaining to the vulnerability of affected populations included in the *Cyclone Fani Joint Situation Analysis* (Needs Assessment Working Group 2019) presented in Chapter 5. Although the methodology proposed in this thesis is exceedingly difficult to verify due to the need for ground-truthing, the study had the advantage of being able to draw upon a spatially referenced survey of households to verify both modelled location information and associated household attributes in a contained geographic area. With the continued proliferation of publicly available microdata and georeferenced survey data, more studies can utilise techniques for localised field validation for similar methodologies moving forward.

Although this thesis only focuses on settled populations, more research can be done to understand population movement. This is particularly important for understanding how populations respond to disaster events. For example, call data records were used in Bangladesh to show movements during and immediately after 2013 Cyclone Mahasen (Lu et al., 2016). In an age where smartphones collect inordinate amounts of data pertaining to movement, spending, and social preferences, the next frontier in population modelling is certain to be dynamic. This is evident in the response to the COVID-19 pandemic, where countries such as Taiwan have tracked quarantined individuals through their mobile home to ensure compliance (Wang et al., 2020). Spatially referenced synthetic households can provide a potentially workable solution for preserving anonymity while also providing a complete set of baseline data useful for further studies linked to spatial agent-based modelling (ABM). Spatial ABM requires a modelled landscape to capture spatial heterogeneities of inputs, which the methodology proposed in this thesis can provide.

Lastly, more must be done to continually evaluate the applicability of technological advancements in geospatial data and analysis relevant to disaster risk management, particularly in developing countries. Those advancements must be integrated into country- and local-level disaster risk management protocols and systems. The key activities proposed by the first priority of the SFDRR, which aims to enhance understanding of disaster risk, emphasise the importance of turning data into actionable information for national and local stakeholders. In the response and recovery phases, existing protocols for data collection efforts often rely on in situ data and direct reports from those situated on the ground. Starting with a higher resolution and more granular baseline understanding of population exposure and vulnerability enables rapid analyses of potentially affected populations when paired with data

detailing hazard extent. By integrating advancements in geospatial analysis, post-disaster needs assessments can help to save lives and improve targeting for effective response and recovery efforts. For mitigation and preparedness efforts, more needs to be done to integrate better baseline population exposure and vulnerability data into investment prioritisation and decision-making, as well as eligibility targeting in disaster-linked social protection programmes and pre-positioning of emergency supplies. Although advancements in high-resolution geospatial analysis have shown significant potential for improving global understanding of disaster risk, more must be done to ensure that those advancements can be practically applied towards improving disaster risk management in developing countries.

Appendix A: Co-author Paper Contribution

Statements

SCHOOL OF GEOGRAHPY AND THE ENVIRONMENT
ENVIRONMENTAL CHANGE INSTITUTE

South Parks Road, Oxford OX1 3QY
Tel: +44(0)7743 508717
steven.rubinyi@chch.ox.ac.uk

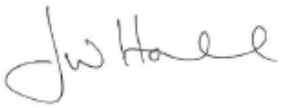


I, Jim W. Hall, certify that Steven Rubinyi completed the majority of the work in the following three Journal Articles below, which form part of his DPhil thesis:

1. *Re-emphasizing the binary mask for high-resolution dasymetric global population modeling*
2. *Spatially referenced synthetic households for high-resolution population modelling*
3. *High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone*

Prof J.W. Hall

Print Name



Signature

Date: 14 April 2020

SCHOOL OF GEOGRAHPY AND THE ENVIRONMENT
ENVIRONMENTAL CHANGE INSTITUTE

South Parks Road, Oxford OX1 3QY
Tel: +44(0)7743 508717
steven.rubinyi@chch.ox.ac.uk

UNIVERSITY OF OXFORD

I, Ran Goldblatt, certify that Steven Rubinyi completed the majority of the work in the following three Journal Articles below, which form part of his DPhil thesis:

1. *High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone*

RAN GOLDBLATT

Print Name

Ran Goldblatt

Signature

04/13/2020

Date

SCHOOL OF GEOGRAHPY AND THE ENVIRONMENT
ENVIRONMENTAL CHANGE INSTITUTE

South Parks Road, Oxford OX1 3QY
Tel: +44(0)7743 508717
steven.rubinyi@chch.ox.ac.uk

UNIVERSITY OF OXFORD

I, Johannes Gussenbauer, certify that Steven Rubinyi completed the majority of the work in the following three Journal Articles below, which form part of his DPhil thesis:

1. *Spatially referenced synthetic households for high-resolution population modelling*
2. *High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone*

JOHANNES GUSSENBAUER
Print Name

Signature

14.04.2020
Date

SCHOOL OF GEOGRAHPY AND THE ENVIRONMENT
ENVIRONMENTAL CHANGE INSTITUTE

South Parks Road, Oxford OX1 3QY
Tel: +44(0)7743 508717
steven.rubinyi@chch.ox.ac.uk

UNIVERSITY OF OXFORD

I, Brian Blankespoor, certify that Steven Rubinyi completed the majority of the work in the following Journal Articles below, which form part of his DPhil thesis:

1. *Re-emphasizing the binary mask for high-resolution dasymetric global population modeling*

Brian BLANKESPOOR
Print Name

Signature

15 APRIL 2020
Date

SCHOOL OF GEOGRAHPY AND THE ENVIRONMENT
ENVIRONMENTAL CHANGE INSTITUTE

South Parks Road, Oxford OX1 3QY
Tel: +44(0)7743 508717
steven.rubinyi@chch.ox.ac.uk

UNIVERSITY OF OXFORD

I, Alexander Kowarik, certify that Steven Rubinyi completed the majority of the work in the following three Journal Articles below, which form part of his DPhil thesis:

1. *High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone*

ALEXANDER KOWARIK
Print Name

Signature

20201014
Date

SCHOOL OF GEOGRAHPY AND THE ENVIRONMENT
ENVIRONMENTAL CHANGE INSTITUTE

South Parks Road, Oxford OX1 3QY
Tel: +44(0)7743 508717
steven.rubinyi@chch.ox.ac.uk

UNIVERSITY OF OXFORD

I, Brad Bottoms, certify that Steven Rubinyi completed the majority of the work in the following three Journal Articles below, which form part of his DPhil thesis:

1. *High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone*

Brad Bottoms

Print Name



Signature

2020 - 04 - 13

Date

SCHOOL OF GEOGRAHPY AND THE ENVIRONMENT
ENVIRONMENTAL CHANGE INSTITUTE

South Parks Road, Oxford OX1 3QY
Tel: +44(0)7743 508717
steven.rubinyi@chch.ox.ac.uk

UNIVERSITY OF OXFORD

I, Jenny Mannix, certify that Steven Rubinyi completed the majority of the work in the following three Journal Articles below, which form part of his DPhil thesis:

1. *High-resolution geospatial analysis for disaster risk management in the Bangladesh Coastal Zone*

Jenny Mannix

Print Name

Jenny Mannix

Signature

4/13/2020

Date

Bibliography

Aamir, M., Pu, Y., Rahman, Z., Tahir, M., Naeem, H., & Dai, Q. (2019). A framework for automatic building detection from low-contrast satellite images. *Symmetry, 11*(1), 3.

Ackerman, E. (2016). Sweep is a $250 LIDAR with range of 40 meters that works outdoors. *IEEE Spectrum, April.*

Aggarwal, A. (2016). Exposure, hazard and risk mapping during a flood event using open source geospatial technology. *Geomatics, Natural Hazards and Risk, 7*(4), 1426-1441.

Alkire, S., & Santos, M. E. (2014). Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index. *World Development, 59*, 251-274.

Alkire, S., Roche, J. M., Ballon, P., Foster, M., Santos, E., & Seth, S. (2015). *Multidimensional poverty measurement and analysis*. USA: Oxford University Press.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., . . . Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv Preprint arXiv:1803.01164.*

Alvarez, M. K., & Cardenas, K. (2019). Evicting slums,'building back better': Resiliency revanchism and disaster risk management in manila. *International Journal of Urban and Regional Research, 43*(2), 227-249.

Amadio, M., Mysiak, J., & Marzi, S. (2019). Mapping socioeconomic exposure for flood risk assessment in italy. *Risk Analysis, 39*(4), 829-845.

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., & Friedl, L. (2017). Earth observation in service of the 2030 agenda for sustainable development. *Geo-Spatial Information Science, 20*(2), 77-96.

Anusha, N., & Bharathi, B. (2019). Flood detection and flood mapping using multi-temporal synthetic aperture radar and optical data. *The Egyptian Journal of Remote Sensing and Space Science.*

Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine, 18*(5), 497-525.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., . . . Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal, 3*, 135-152.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly, 32*(4), 399-418.

Bahia, K., & Suardi, S. (2019). *The state of mobile internet connectivity 2019*. GSM Association.

Baker, J. L. (2012). *Climate change, disaster risk, and the urban poor: Cities building resilience for a changing world* The World Bank.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U., & Nelson, A. (2005). The distribution of people and the dimension of place: Methodologies to improve the global estimation of urban extents. Paper presented at the *International Society*

for Photogrammetry and Remote Sensing, Proceedings of the Urban Remote Sensing Conference, 14-16.

Balk, D. L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S. I., & Nelson, A. (2006). *Determining global population distribution: Methods, applications and data* Elsevier.

Bangladesh Bureau of Statistics. (2012). *2011 bangladesh national census.*

Bangladesh Bureau of Statistics. (2015). Bangladesh population and housing census 2011 (zilla report: Khulna).

Bangladesh Water Development Board. (2012). *Coastal embankment improvement project: Draft final report* Government of the People's Republic of Bangladesh.

Barrett, C. L., Eubank, S., Marathe, A., Marathe, M. V., Pan, Z., & Swarup, S. (2011). Information integration to support model-based policy informatics. *The Innovation Journal: The Public Sector Innovation Journal, 16*(1).

Barthelemy, J., & Toint, P. L. (2013). Synthetic population generation without a sample. *Transportation Science, 47*(2), 266-279.

Baudoin, M., & Wolde-Georgis, T. (2015). Disaster risk reduction efforts in the greater horn of africa. *International Journal of Disaster Risk Science, 6*(1), 49-61.

Beckman, R. J., Baggerly, K. A., & McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice, 30*(6), 415-429.

Belgiu, M., & Stein, A. (2019). Spatiotemporal image fusion in remote sensing. *Remote Sensing, 11*(7), 818.

Bell, R., & Glade, T. (2004). Quantitative risk analysis for landslides? examples from bíldudalur, NW-iceland.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & Von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti. *PLoS Medicine, 8*(8).

Bioresita, F., Puissant, A., Stumpf, A., & Malet, J. (2018). A method for automatic and rapid mapping of water surfaces from sentinel-1 imagery. *Remote Sensing, 10*(2), 217.

Birkin, M. (2019). Spatial data analytics of mobility with consumer data. *Journal of Transport Geography, 76*, 245-253.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete multivariate analysis: Theory and practice* Springer Science & Business Media.

Bobrowsky, P. T. (2013). *Encyclopedia of natural hazards* Springer Dordrecht.

Borgomeo, E., Hall, J. W., & Salehin, M. (2018). Avoiding the water-poverty trap: Insights from a conceptual human-water dynamical model for coastal bangladesh. *International Journal of Water Resources Development, 34*(6), 900-922.

Bostrom, A., French, S. P., & Gottlieb, S. J. (2008). *Risk assessment, modeling and decision support: Strategic directions* Springer Science & Business Media.

Botzen, W. W., Bouwer, L. M., Scussolini, P., Kuik, O., Haasnoot, M., Lawrence, J., & Aerts, J. C. (2019). Integrated disaster risk management and adaptation. *Loss and damage from climate change* (pp. 287-315) Springer.

Brasher, E. A. (2018). Addressing the Failure of Anonymization: Guidance from the European Union's General Data Protection Regulation. *Colum. Bus. L. Rev.*, 209.

Brisco, B., Schmitt, A., Murnaghan, K., Kaya, S., & Roth, A. (2013). SAR polarimetric change detection for flooded vegetation. *International Journal of Digital Earth, 6*(2), 103-114.

Brisco, B. (2015). Mapping and monitoring surface water and wetlands with synthetic aperture radar. *Remote Sensing of Wetlands: Applications and Advances, ,* 119-136.

Bui, D. T., Hoang, N., Martínez-Álvarez, F., Ngo, P. T., Hoa, P. V., Pham, T. D., . . . Costache, R. (2020). A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Science of the Total Environment, 701,* 134413.

Cao, H., Zhang, H., Wang, C., & Zhang, B. (2019). Operational flood detection using sentinel-1 SAR data over large areas. *Water, 11*(4), 786.

Cardona, O. D., Ordaz Schroder, M. G., Reinoso, E., Yamín, L., & Barbat Barbat, H. A. (2010). Comprehensive approach for probabilistic risk assessment (CAPRA): International initiative for disaster risk management effectiveness.

Cassa, C. A., Wieland, S. C., & Mandl, K. D. (2008). Re-identification of home addresses from spatial locations anonymized by gaussian skew. *International Journal of Health Geographics, 7*(1), 45.

Cavallo, E., Powell, A., & Becerra, O. (2010). Estimating the direct economic damages of the earthquake in haiti. *The Economic Journal, 120*(546), F298-F312.

Center for International Earth Science Information Network - CIESIN - Columbia University. (2018). *Gridded population of the world, version 4 (GPWv4): Population density, revision 11* NASA Socioeconomic Data and Applications Center (SEDAC).

Chakraborty, J., Tobin, G. A., & Montz, B. E. (2005). Population evacuation: Assessing spatial variability in geophysical risk and social vulnerability to natural hazards. *Natural Hazards Review, 6*(1), 23-33.

Chen, J., & Zipf, A. (2017). DeepVGI: Deep learning with volunteered geographic information. Paper presented at the *Proceedings of the 26th International Conference on World Wide Web Companion*, 771-772.

Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing, 117*, 11-28.

CIESIN - Columbia University, FAO, & CIAT. (2005). *Gridded population of the world, version 3 (GPWv3): Population count grid, future estimates* NASA Socioeconomic Data and Applications Center (SEDAC) Palisades, NY.

Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010). Framework for analytical quantification of disaster resilience. *Engineering Structures, 32*(11), 3639-3649.

Clark, K. M. (2002). The use of computer modeling in estimating and managing future catastrophe losses. *The Geneva Papers on Risk and Insurance.Issues and Practice, 27*(2), 181-195.

Clement, M. A., Kilsby, C. G., & Moore, P. (2018). Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management, 11*(2), 152-168.

Coppin P, Jonckheere I, Nackaerts K, Muys B, Lambin E. (2004). Digital change detection methods in ecosystem monitoring: A review. international journal of remote sensing 25: 1565-1596.

Costanzo, A., Montuori, A., Silva, J. P., Silvestri, M., Musacchio, M., Doumaz, F., . . . Buongiorno, M. F. (2016). The combined use of airborne remote sensing techniques within a GIS environment for the seismic vulnerability assessment of urban areas: An operational application. *Remote Sensing, 8*(2), 146.

Crichton, D. (1999). The risk triangle. *Natural Disaster Management, 102*, 103.

Croft, W. L., Shi, W., Sack, J., & Corriveau, J. (2017). Comparison of approaches of geographic partitioning for data anonymization. *Journal of Geographical Systems, 19*(3), 221-248.

de la Poterie, Arielle Tozier, & Baudoin, M. (2015). From yokohama to sendai: Approaches to participation in international disaster risk reduction frameworks. *International Journal of Disaster Risk Science, 6*(2), 128-139.

Deichmann, U., Balk, D., & Yetman, G. (2001). Transforming population data for interdisciplinary usages: From census to grid. *Washington (DC): Center for International Earth Science Information Network, 200*(1).

Deichmann, U., Ehrlich, D., Small, C., & Zeug, G. (2011). *Using high resolution satellite data for the identification of urban natural disaster risk.*

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics, 11*(4), 427-444.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., . . . Raska, R. (2018). Deepglobe 2018: A challenge to parse the earth through satellite images. Paper presented at the *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 172-17209.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., . . . Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences, 111*(45), 15888-15893.

DeVries, P. M., Viégas, F., Wattenberg, M., & Meade, B. J. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature, 560*(7720), 632-634.

Dilley, M., Chen, R. S., Deichmann, U., Lerner-Lam, A. L., & Arnold, M. (2005). *Natural disaster hotspots: A global risk analysis* The World Bank.

Diou, C., Lelekas, P., & Delopoulos, A. (2018). Image-based surrogates of socio-economic status in urban neighborhoods using deep multiple instance learning. *Journal of Imaging, 4*(11), 125.

Disaster Management Bureau. (2010). *Standing orders on disaster* Government of the People's Republic of Bangladesh.

Dmowska, A., & Stepinski, T. F. (2017). A high resolution population grid for the conterminous united states: The 2010 edition. *Computers, Environment and Urban Systems, 61*, 13-23.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing, 66*(7), 849-857.

Dolce, M., Hayes, K., Schneider, J. F., De Lannoy, T., Erlich, D., Peter, D., . . . Fujiwara, H. (2012). Global modelling of natural hazard risks—enhancing existing capabilities to address new challenges. Paper presented at the *Organisation for Economic Cooperation and Development (OECD) Global Science Forum, Final Report of the Expert Group.*

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., & Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources, 94,* 87-102.

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., & Baptista, S. R. (2015). Taking advantage of the improved availability of census data: A first look at the gridded population of the world, version 4. *Papers in Applied Geography, 1*(3), 226-234.

Dujardin, S., Jacques, D., Steele, J., & Linard, C. (2020). Mobile phone data for urban climate change adaptation: Reviewing applications, opportunities and key challenges. *Sustainability*, *12*(4), 1501.

Duti, B. M., Hossain, M. M., Khadim, F. K., Rahaman, A. Z., & Quadir, D. A. (2017). Mapping floods in bangladesh using satellite data: An overview. *The Journal of NOAMI, 34*(1).

EERI. (2010). EERI special earthquake report-june 2010: The mw 8.8 chile earthquake of february 27, 2010. *EERI Newsletter.*

Ehrlich, D., Melchiorri, M., Florczyk, A. J., Pesaresi, M., Kemper, T., Corbane, C., . . . Siragusa, A. (2018). Remote sensing derived built-up area and population density to quantify global exposure to five natural hazards over time. *Remote Sensing, 10*(9), 1378.

El Morjani, Zine El Abidine, Ebener, S., Boos, J., Ghaffar, E. A., & Musani, A. (2007). Modelling the spatial distribution of five natural hazards in the context of the WHO/EMRO atlas of disaster risk as a step towards the reduction of the health impact related to disasters. *International Journal of Health Geographics, 6*(1), 8.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro–level estimation of poverty and inequality. *Econometrica, 71*(1), 355-364.

Engstrom, R., Hersh, J., & Newhouse, D. (2017). Poverty from space: Using high-resolution satellite imagery for estimating economic well-being.

Erman, A., Tariverdi, M., Obolensky, M., Chen, X., Vincent, R. C., Malgioglio, S., . . . Yoshida, N. (2019). *Wading out the storm: The role of poverty in exposure, vulnerability and resilience to floods in dar es salaam* The World Bank.

Esch, T., Bachofer, F., Heldens, W., Hirner, A., Marconcini, M., Palacios-Lopez, D., . . . Dech, S. (2018). Where we live—A summary of the achievements and planned evolution of the global urban footprint. *Remote Sensing, 10*(6), 895.

Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., . . . Strano, E. (2017). Breaking new ground in mapping human settlements from space–The global urban footprint. *ISPRS Journal of Photogrammetry and Remote Sensing, 134*, 30-42.

Esch, T., Schenk, A., Ullmann, T., Thiel, M., Roth, A., & Dech, S. (2011). Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and intensity information. *IEEE Transactions on Geoscience and Remote Sensing, 49*(6), 1911-1925.

Facebook Connectivity Lab, & CIESIN - Columbia University. (2016). *High resolution settlement layer (HRSL).*

Filatova, T. (2013). Changing climate, changing behavior: Modeling housing markets responses. *Modeling Social Phenomena in Spatial Context, 2*, 47.

Forget, Y., Shimoni, M., Gilbert, M., & Linard, C. (2018). Complementarity between sentinel-1 and landsat 8 imagery for built-up mapping in sub-saharan africa.

Freire, S., Doxsey-Whitfield, E., MacManus, K., Mills, J., & Pesaresi, M. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution.

Frye, C., & Gilbert, M. (2018). *World population estimated 2016* ESRI.

Garcia, R. A., Oliveira, S. C., & Zêzere, J. L. (2016). Assessing population exposure for landslide risk analysis using dasymetric cartography. *Natural Hazards and Earth System Sciences, 16*(12), 2769.

Gaube, V., Remesch, A., & Smetschka, B. (2016). Planning, residential decisions and energy use in vienna. *Social ecology* (pp. 489-503) Springer.

Gaughan, A. E., Stevens, F. R., Linard, C., Jia, P., & Tatem, A. J. (2013). High resolution population distribution maps for southeast asia in 2010 and 2015. *PloS One, 8*(2).

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences, 114*(50), 13108-13113.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing, 11*(2), 196.

Ghosh, M., & Rao, J. (1994). Small area estimation: An appraisal. *Statistical Science, 9*(1), 55-76.

Ghosh, T., Anderson, S. J., Elvidge, C. D., & Sutton, P. C. (2013). Using nighttime satellite imagery as a proxy measure of human well-being. *Sustainability, 5*(12), 4988-5019.

Gill, J. C., & Malamud, B. D. (2014). Reviewing and visualizing the interactions of natural hazards. *Reviews of Geophysics, 52*(4), 680-722.

Giustarini, L., Hostache, R., Matgen, P., Schumann, G. J., Bates, P. D., & Mason, D. C. (2012). A change detection approach to flood mapping in urban areas using TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing, 51*(4), 2417-2430.

Goda, K., Kiyota, T., Pokhrel, R. M., Chiaro, G., Katagiri, T., Sharma, K., & Wilkinson, S. (2015). *The 2015 gorkha nepal earthquake: Insights from earthquake damage survey* Frontiers.

Goldblatt, R., Jones, N., & Mannix, J. (2020). Assessing OpenStreetMap completeness for management of natural disaster by means of remote sensing: A case study of three small island states (haiti, dominica and st. lucia). *Remote Sensing*, *12*(1), 118.

Gomez, J. (2017). Introduction to the american community survey public use microdata sample (PUMS) files webinar. Retrieved from https://www.census.gov/programs-surveys/acs/news/events/pums-2017.html

Goodchild, M. F., & Lam, N. S. (1980). *Areal interpolation: A variant of the traditional spatial problem* Department of Geography, University of Western Ontario London, ON, Canada.

Gram-Hansen, B. J., Helber, P., Varatharajan, I., Azam, F., Coca-Castro, A., Kopackova, V., & Bilinski, P. (2019). Mapping informal settlements in developing countries using machine learning and low resolution multi-spectral data. Paper presented at the *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 361-368.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., . . . Huse, G. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, *198*(1-2), 115-126.

Grinberger, A. Y., Lichter, M., & Felsenstein, D. (2017). Dynamic agent based simulation of an urban disaster using synthetic big data. *Seeing cities through big data* (pp. 349-382) Springer.

Grossi, P., & Windeler, D. (2005). Sources, nature, and impact of uncertainties on catastrophe modeling. *Catastrophe modeling: A new approach to managing risk* (pp. 69-91) Springer.

Gunasekera, R., Ishizawa, O., Aubrecht, C., Blankespoor, B., Murray, S., Pomonis, A., & Daniell, J. (2015). Developing an adaptive global exposure model to support the generation of country disaster risk profiles. *Earth-Science Reviews, 150*, 594-608.

Habib, A., Shahidullah, M., & Ahmed, D. (2012). The bangladesh cyclone preparedness program. A vital component of the nation's multi-hazard early warning system. *Institutional partnerships in multi-hazard early warning systems* (pp. 29-62). Berlin, Heidelberg: Springer.

Hallegatte, S., Vogt-Schilb, A., Bangalore, M., & Rozenberg, J. (2017). *Unbreakable: Building the resilience of the poor in the face of natural disasters* World Bank Publications.

Hallegatte, S., Rentschler, J., & Walsh, B. (2018). *Building back better: Achieving resilience through stronger, faster, and more inclusive post-disaster reconstruction* World Bank.

Hallegatte, S., Rentschler, J., & Rozenberg, J. (2019). *Lifelines: The resilient infrastructure opportunity* The World Bank.

Haque, U., Hashizume, M., Kolivras, K. N., Overgaard, H. J., Das, B., & Yamamoto, T. (2012). Reduced death rates from cyclones in bangladesh: What more needs to be done? *Bulletin of the World Health Organization, 90*, 150-156.

Harland, K., Heppenstall, A., Smith, D., & Birkin, M. H. (2012). Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques. *Journal of Artificial Societies and Social Simulation*, 15(1).

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., & Borriello, G. (2010). Open data kit: Tools to build information services for developing regions. Paper presented at the *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, 1-12.

Harvey, J. T. (2002). Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, *23*(10), 2071-2095.

Hayes, G. P., Meyers, E. K., Dewey, J. W., Briggs, R. W., Earle, P. S., Benz, H. M., . . . Gold, R. D. (2017). Tectonic summaries of magnitude 7 and greater earthquakes from 2000 to 2015.

Hayhoe, K., Wuebbles, D. J., Easterling, D. R., Fahey, D. W., Doherty, S., Kossin, J. P., . . . Wehner, M. F. (2018). Our changing climate. *Impacts, risks, and adaptation in the united states: Fourth national climate assessment, volume II* (pp. 72-144). Washington, DC, USA: U.S. Global Change Research Program.

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). *Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification* IEEE.

Herman, J. D., Zeff, H. B., Lamontagne, J. R., Reed, P. M., & Characklis, G. W. (2016). Synthetic drought scenario generation to support bottom-up water supply

vulnerability assessments. *Journal of Water Resources Planning and Management, 142*(11), 04016050.

Hewitt, K. (2014). *Regions of risk: A geographical introduction to disasters* Routledge.

Huang, X., Chen, H., & Gong, J. (2018). Angular difference feature extraction for urban scene classification using ZY-3 multi-angle high-resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 135*, 127-141.

Huq, M., Khan, M. F., Pandey, K., Ahmed, M. M. Z., Khan, Z. H., Dasgupta, S., . . . Mukherjee, N. (2010). *Vulnerability of bangladesh to cyclones in a changing climate: Potential damages and adaptation cost* The World Bank.

International Monetary Fund. (2020). Fiscal monitor (april 2020). Retrieved from https://www.imf.org/external/datamapper/datasets/FM

Islam, M. F., Bhattacharya, B., & Popescu, I. (2019). Flood risk assessment due to cyclone-induced dike breaching in coastal areas of bangladesh. *Natural Hazards and Earth System Sciences*, (19), 353-368.

Islam, M. S., Hossain, T. R., Parshi, F. N., Moles, O., Podder, R. K., & Caimi, A. (2015). Disaster resilient rural house designs for different geographic regions of bangladesh. *Proc.of Development and Democracy of Bangladesh: Current Problems and Prospects, UC, Berkeley, USA*,

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790-794.

Ji, S., Wei, S., & Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing, 57*(1), 574-586.

Jia, P., & Gaughan, A. E. (2016). Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in alachua county, florida. *Applied Geography, 66*, 100-108.

Jingai, W., Peijun, S., & Li, Z. (1994). A research on regional distribution of major natural hazards in china [J]. *Acta Geographica Sinica, 1*.

Johnson, B. A., & Iizuka, K. (2016). Integrating OpenStreetMap crowdsourced data and landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the laguna de bay area of the philippines. *Applied Geography, 67*, 140-149.

Jongman, B., Kreibich, H., Apel, H., Barredo, J. I., Bates, P. D., Feyen, L., . . . Ward, P. J. (2012). Comparative flood damage model assessment: Towards a european approach. *Natural Hazards and Earth System Sciences (NHESS), 12*(12), 3733-3752.

Juhász, L., & Hochmair, H. H. (2016). Cross-linkage between mapillary street level photos and OSM edits. *Geospatial data in a changing world* (pp. 141-156) Springer.

Jung, J., & Yun, S. (2020). Evaluation of coherent and incoherent landslide detection methods based on synthetic aperture radar for rapid response: A case study for the 2018 hokkaido landslides. *Remote Sensing, 12*(2), 265.

Kanevski, M., Pozdnukhov, A., & Timonin, V. (2008). Machine learning algorithms for geospatial data. applications and software tools.

Kang, J., Körner, M., Wang, Y., Taubenböck, H., & Zhu, X. X. (2018). Building instance classification using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing, 145*, 44-59.

Kang, Y., Lu, Z., Zhao, C., Zhang, Q., Kim, J., & Niu, Y. (2019). Diagnosis of xinmo (china) landslide based on interferometric synthetic aperture radar observation and modeling. *Remote Sensing, 11*(16), 1846.

Kappes, M. S., Keiler, M., von Elverfeldt, K., & Glade, T. (2012). Challenges of analyzing multi-hazard risk: A review. *Natural Hazards, 64*(2), 1925-1958.

Kelman, I. (2013). Disaster mitigation is cost effective.

Kelman, I. (2015). Climate change and the sendai framework for disaster risk reduction. *International Journal of Disaster Risk Science, 6*(2), 117-127.

Kelman, I. (2017). Linking disaster risk reduction, climate change, and the sustainable development goals. *Disaster Prevention and Management: An International Journal.*

Khan, H., Vasilescu, L. G., & Khan, A. (2008). Disaster management cycle-a theoretical approach. *Journal of Management and Marketing, 6*(1), 43-50.

Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., . . . Irwin, D. (2010). Satellite remote sensing and hydrologic modeling for flood inundation mapping in lake victoria basin: Implications for hydrologic prediction in ungauged basins. *IEEE Transactions on Geoscience and Remote Sensing, 49*(1), 85-95.

Kibria, M. G., & Khan, M. S. A. (2016). Local adaptation practices in response to a super cyclone in the coastal region of bangladesh. *Journal of Modern Science and Technology, 4*(1), 75-85.

Kilic, T., Serajuddin, U., Uematsu, H., & Yoshida, N. (2017). *Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity* The World Bank.

King, A., & Bell, R. (2006). RiskScape new zealand: A multihazard loss modelling tool. Paper presented at the *Proceedings of Earthquake Engineering in the 21st Century (EE-21C) Conference, Topic, 8.*

Kious, W. J., & Tilling, R. I. (1996). *This dynamic earth: The story of plate tectonics* DIANE Publishing.

Kit, O., Lüdeke, M., & Reckien, D. (2012). Texture-based identification of urban slums in hyderabad, india using remote sensing data. *Applied Geography, 32*(2), 660-667.

Klemas, V. (2015). Remote sensing of floods and flood-prone areas: An overview. *Journal of Coastal Research, 31*(4), 1005-1013.

Koks, E. E., Bočkarjova, M., de Moel, H., & Aerts, J. C. (2015). Integrated direct and indirect flood risk modeling: Development and sensitivity analysis. *Risk Analysis, 35*(5), 882-900.

Kolbe, A. R., Hutson, R. A., Shannon, H., Trzcinski, E., Miles, B., Levitz, N., . . . Muggah, R. (2010). Mortality, crime and access to basic needs before and after the haiti earthquake: A random survey of port-au-prince households. *Medicine, Conflict and Survival, 26*(4), 281-297.

Koppel, K., Zalite, K., Voormansik, K., & Jagdhuber, T. (2017). Sensitivity of sentinel-1 backscatter to characteristics of buildings. *International Journal of Remote Sensing*, *38*(22), 6298-6318.

Kovacs, P. (2010). *Reducing the risk of earthquake damage in canada: Lessons from haiti and chile* Institute for Catastrophic Loss Reduction Toronto.

Kramer, A. M., Pulliam, J. T., Alexander, L. W., Park, A. W., Rohani, P., & Drake, J. M. (2016). Spatial spread of the west africa ebola epidemic. *Royal Society Open Science*, *3*(8), 160294.

Kreimer, A., Arnold, M., & Carlin, A. (2003). *Building safer cities: The future of disaster risk* The World Bank.

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6), 455.

Kulp, S., & Strauss, B. H. (2017). Rapid escalation of coastal flood exposure in US municipalities from sea level rise. *Climatic Change*, *142*(3-4), 477-489.

Kulp, S. A., & Strauss, B. H. (2018). CoastalDEM: A global coastal digital elevation model improved from SRTM using a neural network. *Remote Sensing of Environment*, *206*, 231-239.

Lallemant, D., Soden, R., Rubinyi, S., Loos, S., Barns, K., & Bhattacharjee, G. (2017). Post-disaster damage assessments as catalysts for recovery: A look at assessments conducted in the wake of the 2015 gorkha, nepal, earthquake. *Earthquake Spectra*, *33*(1_suppl), 435-451.

Lam, N. S. (1983). Spatial interpolation methods: A review. *The American Cartographer*, *10*(2), 129-150.

Langford, M., & Harvey, J. T. (2001). The use of remotely sensed data for spatial disaggregation of published census population counts. Paper presented at the *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion Over Urban Areas (Cat. no. 01EX482)*, 260-264.

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, *31*(1), 21-26.

Lazo, J. K., Raucher, R. S., Teisberg, T. J., Wagner, C. J., & Weiher, R. F. (2008). Primer on economics for national meteorological and hydrological services. *World Meteorological Organization Voluntary Cooperation Program and National Center for Atmospheric Research, Societal Impacts Program, Boulder, CO.*

Le Cozannet, G., Kervyn, M., Russo, S., Speranza, C. I., Ferrier, P., Foumelis, M., . . . Modaressi, H. (2020). Space-based earth observations for disaster risk management. *Surveys in Geophysics*, 1-27.

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, *89*(10), 93-94.

Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., . . . Frye, C. (2019). The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, *11*(3)

Li, Y., Qi, H., Dai, J., Ji, X., & Wei, Y. (2017). Fully convolutional instance-aware semantic segmentation. Paper presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2359-2367.

Li, G., Zhao, J., Murray, V., Song, C., & Zhang, L. (2019). Gap analysis on open data interconnectivity for disaster risk research. *Geo-Spatial Information Science*, *22*(1), 45-58.

Li, W., He, C., Fang, J., Zheng, J., Fu, H., & Yu, L. (2019). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing*, *11*(4), 403.

Linard, C., & Grippa, T. (2019). *Refining intra-urban population mapping in sub-saharan africa: From land cover to land use*

Linnerooth-Bayer, J., & Hochrainer-Stigler, S. (2015). Financial instruments for disaster risk management and climate change adaptation. *Climatic Change*, *133*(1), 85-100.

Liu, L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, *200*, 170-182.

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. Paper presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8759-8768.

Liverman, D., Moran, E., Rindfuss, R., & Stern, P. (1998). People and pixels. *Linking Remote Sensing and Social Science.National Academy, Washington, DC*,

Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Scientific Data, 4*(1), 1-17.

López-Carresi, A., Fordham, M., Wisner, B., Kelman, I., & Gaillard, J. C. (2013). *Disaster management: International lessons in risk reduction, response and recovery* Routledge.

Lovelace, R., & Dumont, M. (2016). *Spatial microsimulation with R* Chapman and Hall/CRC.

Lu, X., Wrathall, D. J., Sundsøy, P. R., Nadiruzzaman, M., Wetter, E., Iqbal, A., . . . Engø-Monsen, K. (2016). Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in bangladesh. *Global Environmental Change, 38,* 1-7.

Lu, Z., Im, J., Rhee, J., & Hodgson, M. (2014). Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. *Landscape and Urban Planning, 130,* 134-148.

Luongo, G., Perrotta, A., Scarpati, C., De Carolis, E., Patricelli, G., & Ciarallo, A. (2003). Impact of the AD 79 explosive eruption on pompeii, II. causes of death of the inhabitants inferred by stratigraphic analysis and areal distribution of the human casualties. *Journal of Volcanology and Geothermal Research, 126*(3-4), 169-200.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 130,* 277-293.

Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C., & Dow, P. A. (2019). Facebook disaster maps: Aggregate insights for crisis response and recovery. Paper presented at the *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM), Valencia, Spain*, 19-22.

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2016). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing, 55*(2), 645-657.

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. *Earthquake Spectra, 36*(1), 183-208.

Markert, K. N., Chishtie, F., Anderson, E. R., Saah, D., & Griffin, R. E. (2018). On the merging of optical and SAR satellite imagery for surface water mapping applications. *Results in Physics, 9*, 275-277.

Maroko, A., Maantay, J., Pérez Machado, R. P., & Barrozo, L. V. (2019). Improving population mapping and exposure assessment: Three-dimensional dasymetric disaggregation in new york city and são paulo, brazil. *Papers in Applied Geography, 5*(1-2), 45-57.

Martinis, S. (2017). Improving flood mapping in arid areas using sentinel-1 time series data. Paper presented at the *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 193-196.

Martinis, S., Plank, S., & Ćwik, K. (2018). The use of sentinel-1 time-series data to improve flood monitoring in arid areas. *Remote Sensing, 10*(4), 583.

Martinis, S., & Rieke, C. (2015). Backscatter analysis using multi-temporal and multi-frequency SAR data in the context of flood mapping at river saale, germany. *Remote Sensing, 7*(6), 7732-7752.

Mason, D. C., Speck, R., Devereux, B., Schumann, G. J., Neal, J. C., & Bates, P. D. (2009). Flood detection in urban areas using TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing, 48*(2), 882-894.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer, 55*(1), 31-42.

Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass, 3*(2), 727-745.

Mennis, J., & Hultgren, T. (2006). Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science, 33*(3), 179-194.

Merry, K., & Bettinger, P. (2019). Smartphone GPS accuracy study in an urban environment. *PloS One, 14*(7)

Ministry of Water Resources. (2005). *Coastal zone policy.* ().

Miranda, E., Mosqueda, G., Retamales, R., & Pekcan, G. (2012). Performance of nonstructural components during the 27 february 2010 chile earthquake. *Earthquake Spectra, 28*(1_suppl1), 453-471.

Mizutori, M. (2020). Reflections on the sendai framework for disaster risk reduction: Five years since its adoption. *International Journal of Disaster Risk Science,* 1-5.

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood prediction using machine learning models: Literature review. *Water, 10*(11), 1536.

Mou, L., & Zhu, X. X. (2018). Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network. *IEEE Transactions on Geoscience and Remote Sensing, 56*(11), 6699-6711.

Munich Re. (2018). *TOPICS geo: Natural catastrophes 2017*.

Münnich, R., & Schürle, J. (2003). On the simulation of complex universes in the case of applying the german microcensus.

Musa, Z. N., Popescu, I., & Mynett, A. (2015). A review of applications of satellite SAR, optical, altimetry and DEM data for surface water modelling, mapping and parameter estimation. *Hydrology and Earth System Sciences, 19*(9), 3755.

Nagle, N. N., Buttenfield, B. P., Leyk, S., & Spielman, S. (2014). Dasymetric modeling and uncertainty. *Annals of the Association of American Geographers, 104*(1), 80-95.

Nasirzadehdizaji, R., Akyuz, D. E., & Cakir, Z. (2019). Flood mapping and permanent water bodies change detection using sentinel sar data. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences,*

Needs Assessment Working Group. (2019). *Cyclone fani joint situation analysis.* ().

Nieves, J. J., Stevens, F. R., Gaughan, A. E., Linard, C., Sorichetta, A., Hornby, G., . . . Tatem, A. J. (2017). Examining the correlates and drivers of human population distributions across low-and middle-income countries. *Journal of the Royal Society Interface, 14*(137), 20170401.

Nocerino, E., Rodríguez-Gonzálvez, P., & Menna, F. (2019). Introduction to mobile mapping with portable systems. *Laser Scanning: An Emerging Technology in Structural Engineering, 14*, 37.

Olowu, D. (2010). The hyogo framework for action and its implications for disaster management and reduction in africa. *JAMBA: Journal of Disaster Risk Studies, 3*(1), 303-320.

Olson, K. L., Grannis, S. J., & Mandl, K. D. (2006). Privacy protection versus cluster detection in spatial epidemiology. *American Journal of Public Health, 96*(11), 2002-2008.

Othman, S. H., Beydoun, G., & Sugumaran, V. (2014). Development and validation of a disaster management metamodel (DMM). *Information Processing & Management, 50*(2), 235-271.

Paganini, M., Petiteville, I., Ward, S., Dyke, G., Steventon, M., Harry, J., & Kerblat, F. (2018). Satellite earth observations in support of the sustainable development goals. *The CEOS Earth Observation Handbook,*

Paiva, T., Chakraborty, A., Reiter, J., & Gelfand, A. (2014). Imputation of confidential data sets with spatial locations using disease mapping models. *Statistics in Medicine, 33*(11), 1928-1945.

Palacios-Lopez, D., Bachofer, F., Esch, T., Heldens, W., Hirner, A., Marconcini, M., . . . Dech, S. (2019). New perspectives for mapping global population distribution using world settlement footprint products. *Sustainability, 11*(21), 6056.

Papathoma-Köhle, M., Kappes, M., Keiler, M., & Glade, T. (2011). Physical vulnerability assessment for alpine hazards: State of the art and future needs. *Natural Hazards, 58*(2), 645-680.

Peduzzi, P., Dao, H., Herold, C., & Mouton, F. (2009). Assessing global exposure and vulnerability towards natural hazards: The disaster risk index. *Natural Hazards and Earth System Sciences, 9*(4), 1149.

Pekel, J., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature, 540*(7633), 418-422.

Pesaresi, M. (2018). Principles and applications of the global human settlement layer. Paper presented at the *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 2047-2050.

Pielke, R. (2019). Tracking progress on the economic costs of disasters under the indicators of the sustainable development goals. *Environmental Hazards, 18*(1), 1-6.

Poiani, T. H., dos Santos Rocha, R., Degrossi, L. C., & de Albuquerque, J. P. (2016). Potential of collaborative mapping for disaster relief: A case study of OpenStreetMap in the nepal earthquake 2015. Paper presented at the *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 188-197.

Pulvirenti, L., Chini, M., Pierdicca, N., Guerriero, L., & Ferrazzoli, P. (2011). Flood monitoring using multi-temporal COSMO-SkyMed data: Image segmentation and signature interpretation. *Remote Sensing of Environment, 115*(4), 990-1002.

Pulvirenti, L., Chini, M., Marzano, F. S., Pierdicca, N., Mori, S., Guerriero, L., . . . Candela, L. (2012). Detection of floods and heavy rain using cosmo-SkyMed data: The event in northwestern italy of november 2011. Paper presented at the *2012 IEEE International Geoscience and Remote Sensing Symposium*, 3026-3029.

Rahman, A., & Fang, C. (2019). Appraisal of gaps and challenges in sendai framework for disaster risk reduction priority 1 through the lens of science, technology and innovation. *Progress in Disaster Science, 1*, 100006.

Rahman, M. S., & Di, L. (2017). The state of the art of spaceborne remote sensing in flood management. *Natural Hazards, 85*(2), 1223-1248.

Ratcliffe, M., Burd, C., Holder, K., & Fields, A. (2016). Defining rural at the US census bureau. *American Community Survey and Geography Brief*, 1-8.

Ravallion, M. (2011). *On multidimensional indices of poverty* The World Bank.

Rebelo, C., Rodrigues, A. M., & Tenedório, J. A. (2019). Dasymetric mapping using UAV high resolution 3D data within urban areas. *Remote Sensing, 11*(14), 1716.

Reed, F., Gaughan, A., Stevens, F., Yetman, G., Sorichetta, A., & Tatem, A. (2018). Gridded population maps informed by different built settlement products. *Data, 3*(3), 33.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven earth system science. *Nature, 566*(7743), 195-204.

Revilla-Romero, B., Hirpa, F. A., Pozo, J. T., Salamon, P., Brakenridge, R., Pappenberger, F., & De Groeve, T. (2015). On the use of global flood forecasts

and satellite-derived inundation maps for flood monitoring in data-sparse regions. *Remote Sensing, 7*(11), 15702-15728.

Rigaud, K. K., de Sherbinin, A., Jones, B., Bergmann, J., Clement, V., Ober, K., . . . Heuser, S. (2018). Groundswell.

Roberts, M., Blankespoor, B., Deuskar, C., & Stewart, B. (2017). *Urbanization and development: Is latin america and the caribbean different from the rest of the world?* The World Bank.

RTI. (2013). Synthetic household population.

Rubin, D. B. (1993). Statistical disclosure limitation. *Journal of Official Statistics, 9*(2), 461-468.

Sadahiro, Y. (2000). Accuracy of count data transferred through the areal weighting interpolation method. *International Journal of Geographical Information Science, 14*(1), 25-50.

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. *Water Resources Research, 51*(9), 7358-7381.

Schlaffer, S., Matgen, P., Hollaus, M., & Wagner, W. (2015). Flood detection from multi-temporal SAR data using harmonic analysis and change detection. *International Journal of Applied Earth Observation and Geoinformation, 38*, 15-24.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85-117.

Schneider, P. J., & Schauer, B. A. (2006). HAZUS—its development and its future. *Natural Hazards Review, 7*(2), 40-44.

Scholz, C. H. (2019). *The mechanics of earthquakes and faulting* Cambridge university press.

Schumann, G. J. (2015). Preface: Remote sensing in flood monitoring and management. *Remote Sensing*,

Sedda, L., Tatem, A. J., Morley, D. W., Atkinson, P. M., Wardrop, N. A., Pezzulo, C., . . . Rogers, D. J. (2015). Poverty, health and satellite-derived vegetation indices: Their inter-spatial relationship in west africa. *International Health, 7*(2), 99-106.

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., & Dabalen, A. (2015). *Data deprivation: Another deprivation to end* The World Bank.

Ševo, I., & Avramović, A. (2016). Convolutional neural network based automatic object detection on aerial images. *IEEE Geoscience and Remote Sensing Letters, 13*(5), 740-744.

Shaw, R., Mallick, F., & Islam, A. (2013). *Disaster risk reduction approaches in bangladesh* Springer.

Shi, P. (2019). Hazards, disasters, and risks. *Disaster risk science* (pp. 1-48) Springer.

Smajgl, A., & Bohensky, E. (2013). Behaviour and space in agent-based modelling: Poverty patterns in east kalimantan, indonesia. *Environmental Modelling & Software, 45*, 8-14.

Sridharan, H., & Qiu, F. (2013). A spatially disaggregated areal interpolation model using light detection and Ranging-Derived building volumes. *Geographical Analysis, 45*(3), 238-258.

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., . . . Iqbal, A. M. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface, 14*(127), 20160690.

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS One, 10*(2).

Stevens, F. R., Gaughan, A. E., Nieves, J. J., King, A., Sorichetta, A., Linard, C., & Tatem, A. J. (2020). Comparisons of two global built area land cover datasets in methods to disaggregate human population in eleven countries from the global south. *International Journal of Digital Earth, 13*(1), 78-100.

Stockwell, E. G., & Wicks, J. W. (1974). Age heaping in recent national censuses. *Social Biology, 21*(2), 163-167.

Sun, Z., & Müller, D. (2013). A framework for modeling payments for ecosystem services with agent-based models, bayesian belief networks and opinion dynamics models. *Environmental Modelling & Software, 45*, 15-28.

Sunkpho, J., & Ootamakorn, C. (2011). Real-time flood monitoring and warning system. *Songklanakarin Journal of Science & Technology, 33*(2)

Tanaka, K., Fujihara, Y., Hoshikawa, K., & Fujii, H. (2019). Development of a flood water level estimation method using satellite images and a digital elevation

model for the mekong floodplain. *Hydrological Sciences Journal, 64*(2), 241-253.

Tang, Z., Li, Y., Gu, Y., Jiang, W., Xue, Y., Hu, Q., . . . Li, R. (2016). Assessing nebraska playa wetland inundation status during 1985–2015 using landsat data and google earth engine. *Environmental Monitoring and Assessment, 188*(12), 654.

Tapp, A. F. (2010). Areal interpolation and dasymetric mapping methods using local ancillary data sources. *Cartography and Geographic Information Science, 37*(3), 215-228.

Tatem, A. J., Gething, P. W., Pezzulo, C., Weiss, D., & Bhatt, S. (2014). Development of high-resolution gridded poverty surfaces. *See Http://Www.Worldpop.Org.Uk/Resources/Docs/Poverty-Mapping-Report.Pdf,*

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data, 4*(1), 1-4.

Templ, M., Meindl, B., Kowarik, A., & Dupriez, O. (2017). Simulation of synthetic complex data: The R package simPop. *Journal of Statistical Software, 79*(10), 1-38.

Thomson, D. R., Kools, L., & Jochem, W. C. (2018). Linking synthetic populations to household geolocations: a demonstration in Namibia. *Data*, 3(3), 30.

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., . . . Prydz, E. B. (2017). Mapping the world population one building at a time. *arXiv Preprint arXiv:1712.05839,*

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*(367), 519-530.

Tobler, W., Deichmann, U., Gottsegen, J., & Maloy, K. (1997). World population in a grid of spherical quadrilaterals. *International Journal of Population Geography, 3*(3), 203-225.

Toch, E., Lerner, B., Ben-Zion, E., & Ben-Gal, I. (2019). Analyzing large-scale human mobility data: A survey of machine learning methods and applications. *Knowledge and Information Systems, 58*(3), 501-523.

Tom-Aba, D., Olaleye, A., Olayinka, A. T., Nguku, P., Waziri, N., Adewuyi, P., . . . Oguntimehin, O. (2015). Innovative technological approach to ebola virus disease outbreak response in nigeria using the open data kit and form hub technology. *PloS One, 10*(6), e0131000.

Tomaštík, J., Saloň, Š, & Piroh, R. (2016). Horizontal accuracy and applicability of smartphone GNSS positioning in forests. *Forestry: An International Journal of Forest Research, 90*(2), 187-198.

Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science, 19*(4), 385-412.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., . . . Brown, M. (2012). GMES sentinel-1 mission. *Remote Sensing of Environment, 120*, 9-24.

Tracy, M., Cerdá, M., & Keyes, K. M. (2018). Agent-based modeling in public health: Current applications and future directions. *Annual Review of Public Health, 39*, 77-94.

Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., . . . Dutra, E. (2016). The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters, 11*(9), 094014.

Tsyganskaya, V., Martinis, S., Marzahn, P., & Ludwig, R. (2018). Detection of temporary flooded vegetation using sentinel-1 time series data. *Remote Sensing, 10*(8), 1286.

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: A fully automated processing chain. *International Journal of Remote Sensing, 37*(13), 2990-3004.

Uddin, K., Matin, M. A., & Meyer, F. J. (2019). Operational flood mapping using multi-temporal sentinel-1 SAR images: A case study from bangladesh. *Remote Sensing, 11*(13), 1581.

Uddin, M. N., Islam, A. S., Bala, S. K., Islam, G. T., Adhikary, S., Saha, D., ... & Akter, R. (2019). Mapping of climate vulnerability of the coastal region of Bangladesh using principal component analysis. *Applied geography*, 102, 47-57.

UN DESA. (2017). World population prospects 2017.

UN General Assembly. (2015). *Sendai framework for disaster risk reduction 2015–2030.* (). New York: United Nations.

UN General Assembly. (2016). *Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction.* (No. 41). New York: United Nations.

UNISDR. (2004). *Living with risk: A global review of disaster reduction initiatives* United Nations Publications.

UNISDR. (2005). Hyogo framework for action 2005–2015: Building the resilience of nations and communities to disasters. Paper presented at the *Extract from the Final Report of the World Conference on Disaster Reduction (A/CONF. 206/6), , 380*

UNISDR. (2009). UNISDR terminology on disaster risk reduction.

UNISDR. (2015a). *Global assessment report on disaster risk reduction 2015: Making development sustainable: The future of disaster risk management* UN.

UNISDR. (2015b). *The human cost of weather related disasters 1995–2015*.

United States, G. (2019). *Substantial efforts needed to achieve greater progress on high-risk areas.* (). Washington, DC: United States Government Accountability Office.

Utazi, C. E., Wagai, J., Pannell, O., Cutts, F. T., Rhoda, D. A., Ferrari, M. J., . . . Adeniran, A. (2020). Geospatial variation in measles vaccine coverage through routine and campaign strategies in nigeria: Analysis of recent household surveys. *Vaccine*,

Van Dijk, A., & Renzullo, L. J. (2011). Water resource monitoring systems and the role of satellite observations.

Van Niekerk, D. (2014). A critical analysis of the south african disaster management act and policy framework. *Disasters, 38*(4), 858-877.

Vanholder, R., Borniche, D., Claus, S., Correa-Rotter, R., Crestani, R., Ferir, M. C., . . . Portilla, D. (2011). When the earth trembles in the americas: The experience of haiti and chile 2010. *Nephron Clinical Practice, 117*(3), c184-c197.

Voellmy, A., Vrtic, M., Raney, B., Nagel, K., & Axhausen, K. W. (2002). Status of TRANSIMS implementation for switzerland. *Arbeitsberichte Verkehrs-Und Raumplanung, 109.*

Walker, A. J. (1977). An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software (TOMS), 3*(3), 253-256.

Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to COVID-19 in taiwan: Big data analytics, new technology, and proactive testing. *Jama, 323*(14), 1341-1342.

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., . . . Tatem, A. J. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences, 115*(14), 3529-3537.

Weber, E. M., Seaman, V. Y., Stewart, R. N., Bird, T. J., Tatem, A. J., McKee, J. J., . . . Reith, A. E. (2018). Census-independent population mapping in northern nigeria. *Remote Sensing of Environment, 204*, 786-798.

White, L., Brisco, B., Pregitzer, M., Tedford, B., & Boychuk, L. (2014). RADARSAT-2 beam mode selection for surface water and flooded vegetation mapping. *Canadian Journal of Remote Sensing, 40*(2), 135-151.

White, L., Brisco, B., Dabboor, M., Schmitt, A., & Pratt, A. (2015). A collection of SAR methodologies for monitoring wetlands. *Remote Sensing, 7*(6), 7615-7645.

Whitworth, A. (2013). Evaluations and improvements in small area estimation methodologies.

Whitworth, A., Carter, E., Ballas, D., & Moon, G. (2017). Estimating uncertainty in spatial microsimulation approaches to small area estimation: A new approach to solving an old problem. *Computers, Environment and Urban Systems, 63*, 50-57.

Williamson, P., Mitchell, G., & McDonald, A. T. (2002). Domestic water demand forecasting: A static microsimulation approach. *Water and Environment Journal, 16*(4), 243-248.

Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., . . . Hughes, C. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 nepal earthquake. *PLoS Currents, 8.*

Winsemius, H. C., Van Beek, L., Jongman, B., Ward, P. J., & Bouwman, A. (2013). A framework for global river flood risk assessments.

Winsemius, H. C., Jongman, B., Veldkamp, T. I., Hallegatte, S., Bangalore, M., & Ward, P. J. (2015). Disaster risk, climate change, and poverty: Assessing the global exposure of poor people to floods and droughts.

Wohlfart, C., Winkler, K., Wendleder, A., & Roth, A. (2018). TerraSAR-X and wetlands: A review. *Remote Sensing, 10*(6), 916.

World Bank. (2010a). GDP (current US$).

World Bank. (2010b). Poverty maps of bangladesh 2010.

World Bank. (2013). *Bangladesh - first phase of the coastal embankment improvement project (english).* (). Washington DC: World Bank Group.

World Bank. (2014). *Bangladesh - multipurpose disaster shelter project (english).* (). Washington DC:

World Economic Forum. (2020). *The global risks report 2020.* (). Geneva: World Economic Forum.

WorldPop. (2017). Bangladesh 100m population, version 2. Retrieved from https://www.worldpop.org/geodata/summary?id=94

WorldPop, & CIESIN. (2018). *Global high-resolution population denominators project - funded by the bill and melinda gates foundation (OPP1134076)*

Wright, J. K. (1936). A method of mapping densities of population: With cape cod as an example. *Geographical Review*, *26*(1), 103-110.

Wu, S., Qiu, X., & Wang, L. (2005). Population estimation methods in GIS and remote sensing: A review. *GIScience & Remote Sensing*, *42*(1), 80-96.

Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *150*, 59-69.

Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2016). Transfer learning from deep features for remote sensing and poverty mapping. Paper presented at the *Thirtieth AAAI Conference on Artificial Intelligence*,

Xie, Y. (1995). The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems*, *19*(4), 287-306.

Yamazaki, D., O'Loughlin, F., Trigg, M. A., Miller, Z. F., Pavelsky, T. M., & Bates, P. D. (2014). Development of the global width database for large rivers. *Water Resources Research, 50*(4), 3467-3480.

Yuan, K., O'Neil, P., & Torrejon, D. (2020). Landsat's past paves the way for data democratization in earth science. *Data democracy* (pp. 147-161) Elsevier.

Yulianto, F., Sofan, P., Zubaidah, A., Sukowati, K. A. D., Pasaribu, J. M., & Khomarudin, M. R. (2015). Detecting areas affected by flood using multi-temporal ALOS PALSAR remotely sensed data in karawang, west java, indonesia. *Natural Hazards, 77*(2), 959-985.

Zandbergen, P. A. (2011). Dasymetric mapping using high resolution address point datasets. *Transactions in GIS, 15*, 5-27.

Zhao, W., & Du, S. (2016). Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 113*, 155-165.

Zhu, J., Baise, L. G., & Thompson, E. M. (2017). An updated geospatial liquefaction model for global application. *Bulletin of the Seismological Society of America, 107*(3), 1365-1385.

Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., . . . Strobl, P. (2019). Benefits of the free and open landsat data policy. *Remote Sensing of Environment, 224*, 382-385.

Zoraghein, H., & Leyk, S. (2018). Enhancing areal interpolation frameworks through dasymetric refinement to create consistent population estimates across

censuses. *International Journal of Geographical Information Science*, *32*(10), 1948-1976.

Zoraghein, H., & Leyk, S. (2019). Data-enriched interpolation for temporally consistent population compositions. *GIScience & Remote Sensing*, *56*(3), 430-461.