Lancaster University

Modelling spatial processes of infectious diseases

by

James John Chirombo

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

in the

Faculty of Health and Medicine
Lancaster Medical School

September, 2018

To my family

Declaration

This thesis has not been submitted in support of an application for another degree at this or any other university. It is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated. Many of the ideas in this thesis were the product of discussion with my supervisors.

This thesis does not exceed 80,000 words, including footnotes and appendices

James John Chirombo, BSc, MSc
Lancaster, UK
September 2018

Abstract

Human movement plays a key role in the spread of infectious diseases, leading to spatial heterogeneities in disease transmission. An understanding of the causes of these heterogeneities is important in the design, application, and evaluation of public health interventions. In this thesis, we developed a range of statistical models to elucidate spatial dependencies of infection patterns in different populations, and embed existing mobility models within a principled statistical framework. We applied a spatio-temporal generalized linear mixed model to include both climate and non-climate effects on malaria incidence in Malawi while implicitly accounting for spatial dependency and the role of human movement. We further developed methods for real-time assessment of an epidemic by adding spatial information in the calculation of reproductive numbers to account for spatial heterogeneities. A detailed review of mobility models and their use in infectious disease modelling was performed to identify current gaps and opportunities in the field. Finally, a model describing the rate at which human social contact is made in different locations was developed to identify individual-level differences in mobility. The implications for understanding epidemic process and informing control are discussed. With increasing availability of fine-scale mobility data, studying and understanding mobility patterns and their relationship with infectious disease spread will play a key role in developing efficient surveillance and control of emerging and re-emerging diseases.

Acknowledgments

I would like to thank family in Malawi for their support and always having my back.

I am deeply grateful to my supervisors; Dr Jonathan Read, Prof. Peter Diggle and Dr Anja Terlouw for their encouragement and direction during my studies. They were very patient with and put up me through all the ups and downs. The PhD period has been a roll coaster ride and without their continued support, I would not have reached this far. There were always open, and I can confidently say that after about 4 years of study, my supervisors have become colleagues and I look forward to continued collaboration.

I am also thankful to the research staff in CHICAS for always being willing to listen to my questions and taking their precious time to help. Benjamin Taylor, Emanuele Giorgi, Barry Rowlingson, Michael Chipeta, Andreas Parisi and Luigi Sedda.

Many other people have contributed to my studies and I am very thankful. My fellow PhD students in CHICAS at Lancaster University; Olatunji "TJ" Johnson, Eric Chacon, Claudio Frontere, Lisa Koeppel, Anis Ghazali, Poppy Miller and Benjamin Amoah. I could count on their help whenever I needed it. Outside academics, we shared many months of laughter. Those memories will live with me forever.

And lastly, special thanks to Monica for always being a constant source of strength and encouragement throughout this period.

James John Chirombo
Lancaster,
September 2018

Contents

List of Papers

This thesis is based on the following appended papers:

Paper 1: James Chirombo, Pietro Ceccato, Rachel Lowe, Dianne J. Terlouw, Madeleine C. Thomson, Austin Gumbo, Michael Kayenge, Peter J. Diggle & Jonathan Read. *Childhood malaria case incidence in Malawi between 2004 and 2017: Spatio-temporal modelling of climate and non-climate factors.*

Paper 2: James Chirombo, Peter J. Diggle, Diane J. Terlouw & Jonathan M. Read. *Estimation of spatially varying effective reproduction numbers for infectious disease epidemics.*

Paper 3: James Chirombo, Peter J. Diggle, Diane J. Terlouw, & Jonathan Read. *A review of spatial interaction models for human movement patterns affecting the spread of infectious diseases.*

Paper 4: James Chirombo, Peter J. Diggle, Diane J. Terlouw & Jonathan M. Read. *Modelling individual-level differences in human mobility in Southern China.*

Acronyms

ACT	–	Artemisinin-based Combination Therapy
CAR	–	Conditional Autoregressive (model)
CDR	–	Call Detail Record
CHIRPS	–	Climate Hazards Group Infrared Precipitation with Station data
DHIS	–	District Health Information System
DHS	–	Demographic and Health Survey
EVD	–	Ebola Virus Disease
GLM	–	Generalized Linear Model
GLMM	–	Generalized Linear Mixed Model
HMIS	–	Health Management Information System
IRS	–	Indoor Residual Spraying
ITN	–	Insecticide Treated Nets
MERS-COV	–	Middle East Respiratory Syndrome coronavirus (MERS-COV)
MCMC	–	Markov Chain Monte Carlo
MODIS	–	Moderate Resolution Imaging Spectroradiometer
mRDT	–	Malaria Rapid Diagnostic Test
MoH	–	Ministry of Health
MIS	–	Malaria Indicator Survey
NCEP	–	National Centers for Environmental Prediction
NDVI	–	Normalized Difference Vegetative Index
NMCP	–	National Malaria Control Programme
NOAA	–	National Oceanic and Atmospheric Administration
NSO	–	National Statistical Office
RBM	–	Roll Back Malaria
RR	–	Relative Risk
SARS	–	Severe Acute Respiratory Syndrome
SI	–	Serial Interval
SIR	–	Susceptible Infected Recovered
SEIR	–	Susceptible Exposed Infected Recovered
SMR	–	Standardised Morbidity Ratio
WHO	–	World Health Organisation

List of Figures

List of Tables

Part I

Introductory section

Chapter 1

Introduction

1.1 Motivation

The movement of humans from one geographical location to another is a key social characteristic that underpins the spatial extent and spread of many infectious diseases (Read et al. 2012; Sorichetta et al. 2016; Reiner et al. 2014; Riley 2007; Sattenspiel and Lloyd 2009; Soto 2009). For a susceptible person to contract a directly transmitted infectious disease such as influenza, they must first encounter an infectious individual. Mobility, due to social and economic reasons among others, promotes contact rates between susceptible and infected individuals, and can enable the spread of disease to areas with high susceptibility. For vector-borne diseases (VBD) such as malaria and dengue, the role of social connections in driving infections is reduced by the vector movement (Reiner et al. 2014) although it remains an important driver (Stoddard et al. 2009; Cosner et al. 2009). Therefore, the study of and characterization of human mobility behaviour can play a key role in understanding the spread of both direct and indirectly transmitted infectious diseases and, as such, understanding mobility may help improve control. Movement restrictions, such as curfews, quarantine or travel bans, have been used as non-pharmaceutical interventions to control epidemics. These measures essentially aim to limit infectious contacts by reducing human movement. In this thesis, we focus on estimating the mobility of people from a geographical context from three different perspectives: (1) by estimating the pattern of spatial dependency in malaria incidence in Malawi; (2) through the development of an existing method of epidemic assessment to account for and estimate the extent of spatial connectivity of cases; (3) statistical modelling of the spatial dispersal of potentially infectious contacts as measured by a study in China.

1.2 Human mobility, networks, connectivity and disease spread

Mobility can be defined as the movement of people from a place of origin to another destination. It can broadly be categorised into short term and long term movements. In short-term movements, people spend a short amount of time at the destination, usually less than a day, before returning to their place of origin; commuting is another term widely used to refer to these short-term movement patterns. This type of mobility can include movements such as going to the workplace, shopping etc. On the other hand, longer-term movement, where the person making the trip spends a long time at the destination, typically from months and beyond, is better described as migration. Both types of movement can contribute to the geographical spread of a disease, but for many epidemic circumstances the shorter-term movement are thought to dominate disease dynamics. In this thesis, we concentrate on the short-term movements.

Human mobility enhances connectivity between distant communities of people, thereby promoting the spread of infections. Due to different individual characteristics, movement trajectories, which are paths followed by individuals as they move through space over time, may also be unique. For example, some may travel very far to make few contacts, while most may travel only to nearby locations and make many contacts (Read et al. 2014; Brockmann et al. 2006). Despite this variation, human travel patterns can be generalized. It has been observed that humans tend to follow simple reproducible movement patterns, a property that can play a key role in epidemic control (Gonzalez et al. 2008).

The connections between individuals that permit direct transmission or enable indirect transmission can be represented by a network. Various forms of networks have been proposed to describe different forms of human connectivity (Keeling and Eames 2005). Local properties of networks such as the number of contacts are important for disease spread and control. Higher order properties such as clustering, community formation among others can be strongly influential in determining the spread of infection between different parts of the network. However, the network may not necessarily reflect the geographical distribution of individual's home locations. Mobility promotes the long-distance links (edges) between individuals.

1.3 Examples of human movement driving disease spread

Infectious diseases continue to be a threat to human health all over the world (Christian et al. 2013). Both developed and undeveloped countries have their own unique set of infectious disease burdens. The threat of emerging and re-emerging diseases is one of the biggest challenges in today's world (Morens and Fauci 2013; McCloskey et al. 2014). Globalization has made communication, whether by air, road or rail easy such that any one area of the world is well connected to another (Tatem et al. 2006). This means that populations of people are always in touch with outside populations and therefore, directly transmissible infectious diseases, not native to a location can easily be passed on to another population through human interaction.

Historically, human mobility through long range trade, has been found to be responsible for driving the spread of diseases such as plaque in pre-industrial Europe (Yue et al. 2017). In recent years, severe acute respiratory syndrome (SARS) was the first major emerging disease outbreak in this millennium and demonstrates the importance of long-range human mobility in spreading infections between countries. First discovered in Southern China in 2002-2003, the epidemic soon went global with cases reported as far as Canada within a relatively short period of time (Parashar and Anderson 2004; Svoboda et al. 2004). The recent 2013-16 West Africa Ebola virus disease (EVD) outbreak also spread beyond West Africa to Europe and the United States, though it did not lead into an outbreak in these countries (WHO 2016). The Middle East Respiratory Syndrome coronavirus (MERS-COV) was first identified in Saudi Arabia but spread to over 25 countries including South Korea where it caused a large nosocomial outbreak. The disease spread to South Korea through a returning individual who had visited the Middle East (Hui et al. 2015). Risk of further spread from the Middle East was particularly high due to the presence of large air transport hubs and annual gatherings of millions of people from around the world during the Haj and Umrah pilgrimage (Gardner et al. 2016).

The Zika epidemic also rapidly spread across countries mainly with air travel playing a prominent role in this rapid inter-country transmission (Bogoch et al. 2016). In South East Asia, Tian et al. (2017) found the growth in incidence of Dengue virus serotypes 1,2 and 3 to be associated with air travel. In addition, travel and migration also plays a role in the increasing levels of antimicrobial resistance (Bogoch et al. 2016).

For vector-borne infectious diseases, movement of people enables the disease-transmitting agents to go further than they would naturally do (Tatem et al. 2006). In the case of malaria, for example, mosquitoes which act as vectors can

be introduced into an area that previously did not support mosquitoes hence introducing or re-introducing malaria into the area. Mosquitoes themselves have a limited range of migration (Kaufmann and Briegel 2004), thus human mobility plays a key role in malaria transmission.

1.4 Characteristics of human movement

Human mobility exhibits spatial, temporal and connectivity characteristics (Karamshuk et al. 2011). The spatial component captures characteristics such as distance travelled. The temporal component pertains to how mobility behaviour varies with time, such as time spent at a particular location. Lastly, connectivity refers to the interactions that may occur between individuals in a location, including between those normally resident and travellers. In terms of distance covered, the majority of the population do not travel very far from their origin. A recent study has found that most social contacts in China are within 1km of a person's home with some occurring over 500km away (Read et al. 2014). The distances travelled by individuals have been found to be well approximated by a power law distribution, $p(\Delta r) \sim (\Delta r)^{-(1+\beta)}$, where Δr is the travel distance and the exponent $\beta < 2$ (Brockmann et al. 2006). This observation for the distance travelled is supported by spatial interaction models such as the gravity model which show reduction in interaction as distance between location pairs increases. This distribution describes the general observation that most of the time, people make short journeys with a few occasional long journeys. For the temporal characteristics, the time spent making contacts at a destination has also been studied (Tilahun and Levinson 2017) and it has been shown to follow a power law distribution (Song et al. 2010).

One of the key challenges in applying the power law in modelling distance travelled is that it is rarely known whether an observed quantity is indeed from a power law distribution (Clauset et al. 2009; Hanel et al. 2017). Estimating the parameter β is also a difficult task (Bauke 2007). Furthermore, the decision that a quantity of interest is drawn from a power law should be backed up by robust tests. In Brockmann et al. (2006), no formal tests were carried out to ascertain whether the observed values came from a proper power law distribution. Clauset et al. (2009) developed methods for parameter estimation and determining if the data comes from a power law: after testing on datasets previously categorized as power laws, some were ruled out to have come from this distribution.

1.5 Implications for disease control

Networks possess two key characteristics of resilience and fragility. The resilience is shown in the sense that when a particular location in the network, with few edges (i.e. low edge to node ratio), is disrupted, the network does not break down and continues to function (Gao et al. 2016). On the other hand, when a single node with so many connections is taken out, the whole network is disrupted. This knowledge can play a key role in the control of infectious diseases by informing the targeting of interventions. Human interaction is known to follow a scale-free distribution. In a sexually transmitted disease such as HIV/AIDS, targeting individuals with multiple sex partners can lead to quicker results in stopping the epidemic. In the same way, there exist super-spreaders of diseases, those with multiple infectious contacts with susceptible people (Stein 2011). These individuals are responsible for the rapid spread of a disease in a susceptible population. Better interventions that target these superspreaders can more quickly bring epidemics under control.

1.6 Types of human movement

1.6.1 Individual and mass movement

In most applications of human mobility, movements are aggregated at the spatial unit level. This mass movement from one location to another is known as flux and these movement patterns can be described in terms of a Markov chain. The places visited by the moving person make up the state space of the Markov chain. If the population-scale probabilities of an individual moving from one place to another are known, a transitional matrix for movement between locations can be constructed. An assumption that the probability of a person moving to a location j depends on their current location i only and not where they have been in the past ($i, j = A, B, C, D; i \neq j$) is made. Figure 1.1 represents a simple Markov chain representation for commuting flows between 4 hypothetical locations A, B, C and D.

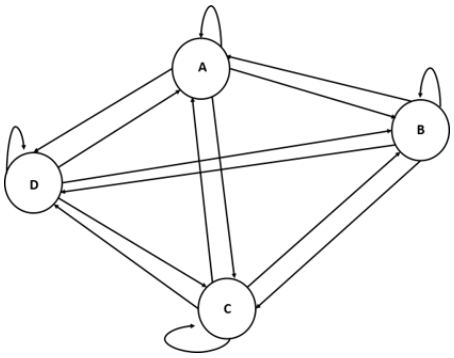

Figure 1.1: Markov chain representation of commuting flow between four hypothetical locations

The movement is represented as a graph where the nodes (i.e. the cities) are the state spaces and the edges are the journeys between locations. Of note here are the individuals that move around within their current locations; they do not make any journeys between locations.

At the aggregate level, the sizes of populations in different locations may play a key role in the movement of people (Congdon 2010). Some locations may have a higher attractive power to 'pull' people from the surrounding areas due to the presence of more opportunities for people, such as jobs, school and shopping. These mass movements between locations are described by the gravity and radiation models which will be discussed in detail in Chapter 4.

1.7 Modelling approaches for infectious diseases

Different mobility patterns could lead to different interaction patterns between individuals. Social behaviours governing both mobility and interaction patterns, therefore, can determine the landscape of the contact network upon which infections may spread. For directly transmitted infectious diseases, these behaviours can ultimately determine how rapidly the disease will spread. Sexually transmitted diseases, for example, are determined by the connectivity and network structure that is formed between sexual partners. In order to account for all the possible spatial heterogeneities in disease transmission, modelling approaches, whether statistical or process based are used.

1.7.1 Statistical models

Statistical models consider the relationships between the outcome of interest and a set of covariates. Generalised linear models (GLMs) (McCullagh and Nelder 1989) play a central role in modelling this relationship which can be represented by the

functional relationship

$$y = f(x, \Theta) + \varepsilon, \tag{1.1}$$

where y is the response variable, x is a set of covariates, Θ is a set of corresponding parameters and ε is the error term. The GLM can be extended to include unobserved spatial effects by the inclusion of random effect terms into the standard GLM. The random effects terms capture the unobserved characteristics of the population that are thought to contribute substantially to the response variable and hence lead to better estimates. Human mobility patterns, in situations where explicit information is not available to include in the model as a covariate, can be captured by the random effects. This is a key strength of statistical models over process based models.

1.7.2 Process based models

Process-based models can handle the network structure, human mobility patterns and connectivity behaviour. Example include network-based metapopulation models which are widely used (Eames and Keeling 2002; Eubank 2005). These models explicitly describe how transmission of an infection occurs. The key components of the transmission processes have to be quantified and represented in the model. Therefore, they are relatively harder to fit to the data compared to statistical models. If all the key process parameters are available, these models can have very high predictive power as they are closer to the real-life situation. However, these parameters are rarely known when fitting the model. Another challenge for process-based models is validation. Usually, there are not enough data for different scenarios for validation. Compared to statistical models, process-based models are usually fitted to small datasets.

1.8 Statistical modelling of spatial processes of infectious diseases

Two of the main statistical modelling approaches for point processes are discussed in this thesis. In the first case, the points are aggregated at discrete spatial locations of interest thereby yielding counts. These are known as areal data and inference is based on these spatial units. Counts take on non-negative integer values which are better described by the Poisson distribution hence a Poisson GLM is typically used to model the contribution of other covariates to the observed counts. When the variance and mean are not the same, a phenomenon known as over-dispersion, negative binomial regression can instead be used. Disease mapping

studies commonly use these regression approaches to estimate disease risk in a spatial region. Examples include (Waller et al. 1997; Kelsall and Wakefield 2002).

In some settings, the data are not aggregated, and it is assumed that they are continuous in space rather than discrete. Point process approaches are the models of choice in this case. Depending on the application, either purely temporal or spatial point process are observed. Sometimes it is necessary to model the joint spatio-temporal process. In the spatio-temporal case, the rate at which points (events of interest such as new disease cases) occur in a spatial region within a specified time window is modelled. Examples of spatio-temporal point process models include (Diggle 2006; Diggle et al. 2010).

In this thesis, we apply both modelling approaches to areal and point process data with applications to infectious diseases. We apply the former to estimate and map malaria risk at the district level in Malawi and the spatio-temporal point process methods for evaluating infectious disease spread and understanding human social contact patterns which are crucial for the propagation of directly transmissible infectious diseases.

1.8.1 Mapping disease incidence in Malawi

For common VBD such as malaria, understanding the relationship between climate, socio-economic factors among others with malaria incidence is important from a public health perspective. A special interest in many malaria-endemic countries today is on understanding the interactions between malaria (and other VBD such as dengue) with climate change. The World Health Organization (WHO) estimates that more than half of the world's population is currently at risk of climate-sensitive VBD with over one million deaths every year (Campbell-Lendrum et al. 2015). It is further estimated that a $1.0 - 3.5°C$ increase in average global temperatures by the year 2100 will increase the likelihood of many VBD in new areas (Githeko et al. 2000).

It is worth remembering that statistical models do not explain all the variation in the outcome. For example, modelling malaria incidence over time and space in terms of climate will not be able to attribute all the incidence pattern to climate changes. The addition of other covariates may improve model fit and prediction. Hence we may add non-climatic covariates such as socioeconomic factors. Sometimes, information necessary to improve the model may not be available forcing modellers to use proxies. In such cases, the model will be compromised due to lack of information, leading to big error component. In a spatio-temporal model for malaria and climate interactions, we might improve the model by adding non-climate confounders, if available.

One piece of information that is likely to contribute to VBD spread is human mobility (Stoddard et al. 2009). Detailed information on mobility patterns is difficult to find in many settings, especially in the developing world. Near real-time mobility data is even harder to find, thus limiting modelling to low-resolution temporal and spatial scales. Therefore, a crucial driver of VBD is not usually included in models due to its scarcity. One solution to this problem is to use random effects to capture some unobserved characteristics of the population. An assumption that these will cover several effects including the role of mobility can be made, though this effect cannot be isolated on its own in the absence of observed data. Human mobility, though a useful predictor, is not included in our models in Chapter 2. We will make an assumption that its effect is captured by the inclusion of spatially structured random effects.

1.8.2 Modelling infectious disease potential

To understand disease spread, a metric known as the effective reproduction number (R_t) is used to quantify the transmission potential. R_t is defined as the average number of secondary cases that an infected individual will infect by the end of their infectious period. If $R_t > 1$, it means the epidemic is growing while $R_t < 1$ implies the epidemic is getting under control and will eventually be contained. The R_t only tells of how the infection is progressing over time. While this is useful, spatial heterogeneities in disease spread are not integrated in the R_t estimate. The R_t can yield unreliable results in locations with high levels of heterogeneity.

In this thesis, we extend the method for R_t estimation to capture this spatial heterogeneity. The point process approach discussed above is used to describe the rate at which new cases develop in a location because of possible contact with infectious individuals from the same or another location. This spatio-temporal point process approach is well placed to capture the spatial heterogeneity in disease spread.

1.9 Outline of the thesis

This thesis is organised as a compilation of papers: In Chapter 2, we present the spatio-temporal statistical model for modelling climate and non-climate drivers of malaria in Malawi. This chapter sets the scene for the importance of spatial processes in infectious diseases, and human mobility as a possible catalyst for infectious disease spread. In Chapter 3, we introduce a new method for estimating spatial-temporal reproductive numbers as a way of improving upon the current methods which do not take into consideration the spatial heterogeneity in disease

transmission. We apply the method to the West Africa Ebola outbreak of 2013-16. Chapter 4 is a detailed review of human mobility models, their history and their application in infectious disease modelling. Chapter 5 introduces a model for individual-level differences in movement patterns. These individual-level movements have implications in the spread of diseases such as influenza. A case study of geographically-located human interaction patterns in Guangdong province of China is used as an application. Chapter 6 discusses the results of the thesis in the context of all the component papers/chapters. It also discusses the contributions of the thesis and explores areas of future work in the field of human mobility modelling from the perspective of infectious disease epidemiology.

References

Bauke, Heiko (2007). "Parameter estimation for power-law distributions by maximum likelihood methods." In: *European Physical Journal B–Condensed Matter* 58.2.

Bogoch, Isaac I, Oliver J Brady, Moritz UG Kraemer, Matthew German, Marisa I Creatore, Manisha A Kulkarni, John S Brownstein, Sumiko R Mekaru, Simon I Hay, Emily Groot, et al. (2016). "Anticipating the international spread of Zika virus from Brazil." In: *The Lancet* 387.10016, pp. 335–336.

Brockmann, Dirk, Lars Hufnagel, and Theo Geisel (2006). "The scaling laws of human travel." In: *Nature* 439.7075, pp. 462–465.

Campbell-Lendrum, Diarmid, Lucien Manga, Magaran Bagayoko, and Johannes Sommerfeld (2015). "Climate change and vector-borne diseases: what are the implications for public health research and policy?" In: *Phil. Trans. R. Soc. B* 370.1665, p. 20130552.

Christian, Kira A, Kashef Ijaz, Scott F Dowell, Catherine C Chow, Rohit A Chitale, Joseph S Bresee, Eric Mintz, Mark A Pallansch, Steven Wassilak, Eugene McCray, et al. (2013). "What we are watching—five top global infectious disease threats, 2012: a perspective from CDC's Global Disease Detection Operations Center." In: *Emerging health threats journal* 6.1, p. 20632.

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman (2009). "Power-law distributions in empirical data." In: *SIAM review* 51.4, pp. 661–703.

Congdon, Peter (2010). "Random-effects models for migration attractivity and retentivity: a Bayesian methodology." In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173.4, pp. 755–774.

Cosner, C, JC Beier, RS Cantrell, D Impoinvil, L Kapitanski, MD Potts, A Troyo, and S Ruan (2009). "The effects of human movement on the persistence of vector-borne diseases." In: *Journal of theoretical biology* 258.4, pp. 550–560.

Diggle, Peter J (2006). "Spatio-temporal point processes, partial likelihood, foot and mouth disease." In: *Statistical methods in medical research* 15.4, pp. 325–336.

Diggle, Peter J, Irene Kaimi, and Rosa Abellana (2010). "Partial-Likelihood Analysis of Spatio-Temporal Point-Process Data." In: *Biometrics* 66.2, pp. 347–354.

Eames, Ken TD and Matt J Keeling (2002). "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases." In: *Proceedings of the National Academy of Sciences* 99.20, pp. 13330–13335.

Eubank, Stephen (2005). "Network based models of infectious disease spread." In: *Japanese journal of infectious diseases* 58.6, S.

Gao, Jianxi, Baruch Barzel, and Albert-László Barabási (2016). "Universal resilience patterns in complex networks." In: *Nature* 530.7590, p. 307.

Gardner, Lauren M, Abrar A Chughtai, and C Raina MacIntyre (2016). "Risk of global spread of Middle East respiratory syndrome coronavirus (MERS-CoV) via the air transport network." In: *Journal of travel medicine* 23.6.

Githeko, Andrew K, Steve W Lindsay, Ulisses E Confalonieri, and Jonathan A Patz (2000). "Climate change and vector-borne diseases: a regional analysis." In: *Bulletin of the World Health Organization* 78.9, pp. 1136–1147.

Gonzalez, Marta C, Cesar A Hidalgo, and Albert-Laszlo Barabasi (2008). "Understanding individual human mobility patterns." In: *Nature* 453.7196, pp. 779–782.

Hanel, Rudolf, Bernat Corominas-Murtra, Bo Liu, and Stefan Thurner (2017). "Fitting power-laws in empirical data with estimators that work for all exponents." In: *PloS one* 12.2, e0170920.

Hui, David S, Stanley Perlman, and Alimuddin Zumla (2015). "Spread of MERS to South Korea and China." In: *The lancet Respiratory medicine* 3.7, pp. 509–510.

Karamshuk, Dmytro, Chiara Boldrini, Marco Conti, and Andrea Passarella (2011). "Human mobility models for opportunistic networks." In: *IEEE Communications Magazine* 49.12, pp. 157–165.

Kaufmann, Christian and Hans Briegel (2004). "Flight performance of the malaria vectors Anopheles gambiae and Anopheles atroparvus." In: *Journal of vector ecology* 29, pp. 140–153.

Keeling, Matt J and Ken TD Eames (2005). "Networks and epidemic models." In: *Journal of the Royal Society Interface* 2.4, pp. 295–307.

Kelsall, Julia and Jonathan Wakefield (2002). "Modeling spatial variation in disease risk: a geostatistical approach." In: *Journal of the American Statistical Association* 97.459, pp. 692–701.

McCloskey, Brian, Osman Dar, Alimuddin Zumla, and David L Heymann (2014). "Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread." In: *The Lancet infectious diseases* 14.10, pp. 1001–1010.

McCullagh, Peter and John A Nelder (1989). *Generalized linear models*. Vol. 37. CRC press.

Morens, David M and Anthony S Fauci (2013). "Emerging infectious diseases: threats to human health and global stability." In: *PLoS pathogens* 9.7, e1003467.

Parashar, Umesh D and Larry J Anderson (2004). *Severe acute respiratory syndrome: review and lessons of the 2003 outbreak*.

Read, JM, WJ Edmunds, S Riley, J Lessler, and DAT Cummings (2012). "Close encounters of the infectious kind: methods to measure social mixing behaviour." In: *Epidemiology & Infection* 140.12, pp. 2117–2130.

Read, Jonathan M, Justin Lessler, Steven Riley, Shuying Wang, Li Jiu Tan, Kin On Kwok, Yi Guan, Chao Qiang Jiang, and Derek AT Cummings (2014). "Social

mixing patterns in rural and urban areas of southern China." In: *Proceedings of the Royal Society of London B: Biological Sciences* 281.1785, p. 20140268.

Reiner, Robert C, Steven T Stoddard, and Thomas W Scott (2014). "Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal." In: *Epidemics* 6, pp. 30–36.

Riley, Steven (2007). "Large-scale spatial-transmission models of infectious disease." In: *Science* 316.5829, pp. 1298–1301.

Sattenspiel, Lisa and Alun Lloyd (2009). *The geographic spread of infectious diseases: models and applications*. Princeton University Press.

Song, Chaoming, Tal Koren, Pu Wang, and Albert-László Barabási (2010). "Modelling the scaling properties of human mobility." In: *Nature Physics* 6.10, pp. 818–823.

Sorichetta, Alessandro, Tom J Bird, Nick W Ruktanonchai, Elisabeth zu Erbach-Schoenberg, Carla Pezzulo, Natalia Tejedor, Ian C Waldock, Jason D Sadler, Andres J Garcia, Luigi Sedda, et al. (2016). "Mapping internal connectivity through human migration in malaria endemic countries." In: *Scientific data* 3, p. 160066.

Soto, SM (2009). "Human migration and infectious diseases." In: *Clinical Microbiology and Infection* 15.s1, pp. 26–28.

Stein, Richard A (2011). "Super-spreaders in infectious diseases." In: *International Journal of Infectious Diseases* 15.8, e510–e513.

Stoddard, Steven T, Amy C Morrison, Gonzalo M Vazquez-Prokopec, Valerie Paz Soldan, Tadeusz J Kochel, Uriel Kitron, John P Elder, and Thomas W Scott (2009). "The role of human movement in the transmission of vector-borne pathogens." In: *PLoS neglected tropical diseases* 3.7, e481.

Svoboda, Tomislav, Bonnie Henry, Leslie Shulman, Erin Kennedy, Elizabeth Rea, Wil Ng, Tamara Wallington, Barbara Yaffe, Effie Gournis, Elisa Vicencio, et al. (2004). "Public health measures to control the spread of the severe acute respiratory syndrome during the outbreak in Toronto." In: *New England Journal of Medicine* 350.23, pp. 2352–2361.

Tatem, Andrew J, David J Rogers, and SI Hay (2006). "Global transport networks and infectious disease spread." In: *Advances in parasitology* 62, pp. 293–343.

Tian, Huaiyu, Zhe Sun, Nuno Rodrigues Faria, Jing Yang, Bernard Cazelles, Shanqian Huang, Bo Xu, Qiqi Yang, Oliver G Pybus, and Bing Xu (2017). "Increasing airline travel may facilitate co-circulation of multiple dengue virus serotypes in Asia." In: *PLoS neglected tropical diseases* 11.8, e0005694.

Tilahun, Nebiyou and David Levinson (2017). "Contacts and meetings: Location, duration and distance traveled." In: *Travel Behaviour and Society* 6, pp. 64–74.

Waller, Lance A, Bradley P Carlin, Hong Xia, and Alan E Gelfand (1997). "Hierarchical spatio-temporal mapping of disease rates." In: *Journal of the American Statistical association* 92.438, pp. 607–617.

WHO, World Health Organization Ebola Response Team (2016). "After Ebola in West Africa—unpredictable risks, preventable epidemics." In: *New England Journal of Medicine* 375.6, pp. 587–596.

Yue, Ricci PH, Harry F Lee, and Connor YH Wu (2017). "Trade routes and plague transmission in pre-industrial Europe." In: *Scientific reports* 7.1, p. 12973.

Part II

Appended papers

Chapter 2

Childhood malaria case incidence in Malawi between 2004 and 2017: Spatio-temporal modelling of climate and non-climate factors

Chirombo J, Ceccato P, Lowe R, Terlouw D.J., Thomson M.C, Gumbo A,Kayenge M, Diggle P.J., Read J.M.

2. Childhood malaria case incidence in Malawi between 2004 and 2017: Spatio-temporal modelling of climate and non-climate factors

Chirombo J, Ceccato P, Lowe R, Terlouw D.J., Thomson M.C,
Gumbo A,Kayenge M, Diggle P.J., Read J.M.

Abstract

Background: Malaria transmission is influenced by a complex interplay of factors including climate, socio-economic, environmental factors and interventions. Malaria control efforts across Africa have shown a mixed impact. Climate driven factors may play an increasing role with climate change. Efforts to strengthen routine facility-based monthly malaria data collection across Africa create an increasingly valuable data source to interpret burden trends and monitor control programme progress. A better understanding of the association with other climatic and non-climatic drivers of malaria incidence over time and space may help guide and interpret the impact of interventions.

Methods: Routine monthly paediatric outpatient clinical malaria case data were compiled from 27 districts in Malawi between 2004 and 2017, and analysed in combination with data on climatic, environmental, socio-economic and interventional factors and district level population estimates. A spatio-temporal generalized linear mixed model was fitted using Bayesian inference, in order to quantify the strength of association of the various risk factors with district-level variation in clinical malaria rates in Malawi, and visualised using maps.

Results: Between 2004 and 2017 reported childhood clinical malaria case rates showed a slight increase, from 50 to 53 cases per 1000 population, with considerable variation across the country between climatic zones. Climatic and environmental factors, including average monthly air temperature and rainfall anomalies, normalized difference vegetative index (NDVI) and RDT use for diagnosis showed a significant relationship with malaria incidence. Temperature in the current month and in each of the 3 months prior showed a significant relationship with the disease incidence unlike rainfall anomaly which was associated with malaria incidence at only three months prior. Estimated risk maps show relatively high risk along the lake and Shire valley regions of Malawi.

Conclusion: Our modelling approach can identify locations likely to have unusually

high or low risk of malaria incidence across Malawi, and distinguishes between contributions to risk that can be explained by measured risk-factors and unexplained residual spatial variation. Also, spatial statistical methods applied to readily available routine data provides an alternative information source that can supplement survey data in policy development and implementation to direct surveillance and intervention efforts.

1 Introduction

While malaria has declined across Africa, analyses exploring the impact of nationally implemented control interventions have shown a mixed impact, with recent analyses of malaria prevalence data across Africa since 1900 showing a complex range of driving factors including climate, socio-economic and environmental factors that may all depend on time and local context (Snow et al. 2017). Climate affects many aspects of the transmission dynamics of malaria by its effects on the vector biology (Parham and Michael 2010; Wu et al. 2016; Altizer et al. 2013; Githeko et al. 2000; Cash et al. 2013), and is expected to play an increasing role with progressive climate change. While global malaria control progress is monitored through malaria prevalence estimates from household surveys, national programmes in endemic countries often use facility based data to set impact targets and monitor progress, as this data is available on an ongoing basis and relates to disease burden rather than transmission. With efforts to strengthen the quality of routine facility-based monthly malaria data collection across Africa and progress in analytical methods to analyse collated data from multiple sources, this becomes an increasingly important data source.

Vector population larvae development depends on sufficient rainfall, yet excess rainfall can reduce numbers due to excessive water flow (Paaijmans et al. 2007). Temperature also plays a crucial role as the main vectors, such as tropical Anopheles mosquitoes, require temperatures between 16°C and 32°C to complete their life cycles. At higher and lower temperatures, there is high mosquito mortality (Lafferty 2009). Consequently, malaria displays seasonal patterns in response to changing climatic conditions.

In addition to climate, socio-economic factors play a critical role in malaria transmission (Yadav et al. 2014; Nkegbe et al. 2017). Therefore, addressing malaria through the design of optimal interventions can benefit from a clear understanding of the impact of both climate and non-climate factors.

Use of climate data to improve our understanding of the observed trends and patterns in climate-sensitive diseases has not been widely undertaken in many African countries due to incomplete or unreliable climate and disease incidence

data. The use of climate data derived from remote sensing provides an opportunity to investigate the impact of climate on malaria, even for areas where climate data from weather stations are sparse or non-existent.

Statistical models for aggregate and point-level data have been used to improve our understanding of the interactions between vector-borne diseases (VBD) and environmental conditions (Diggle et al. 2002; Lowe et al. 2011; Kazembe et al. 2006). Furthermore, work has been carried out towards the development of early warning systems for VBD such as malaria and dengue (Lowe et al. 2011; Lowe et al. 2014; Thomson et al. 2005; Thomson et al. 2006; Connor and Mantilla 2008). In many settings, however, non-climatic conditions also play a key role in driving VBD and these act as confounding factors (Tompkins et al. 2016; Lindblade et al. 2000; Bødker et al. 2000). A purely climate-based model may thus not be sufficient to capture the complex relationships between VBD and the total environmental in general (Stewart-Ibarra and Lowe 2013).

We previously explored the roles of climate, geographic and socio-economic factors on malaria in Malawi and mapped disease incidence for the period 2004-2011 (Lowe et al. 2013). Since then, national control efforts have scaled up substantially, including the successful scale up of effective artemisinin-based combination therapy (ACT) since 2009, of malaria rapid diagnostic tests (RDTs) since 2011, and use of regular national net distribution campaigns since 2012 to move towards universal net coverage. Overall, malaria prevalence in children below 5 years of age has declined from 43% in 2010 to 24% in 2017 (NMCP 2011; NMCP and ICF 2018).

This study aims to add to the evidence on the linkages between climate and malaria in Malawi and shows how the contribution of relevant non-climatic confounding factors can be visualised in a way that may help inform national malaria control programmes on options to take those factors into account and mitigate the impact of climate change. Using age-stratified malaria data from Malawi with climatic and non-climatic covariates we built a spatio-temporal statistical model implemented in a Bayesian inferential framework and mapped explained and unexplained components of the spatio-temporal variation in malaria incidence.

2 Methods

Malawi context

Malaria is endemic to Malawi but with spatially varying levels of transmission (Townes et al. 2013), across a varied geographical landscape, from lowlands to highlands. Lakeshore districts generally have higher malaria prevalence than other districts. The country is divided into 5 climatic zones by the government's

meteorology department across 28 districts. Districts along the lake are generally of low altitude and have high average temperatures with average elevation ranges from 500m above sea level along the lake and Shire valley to over 1500 in the central areas. Rainfall across Malawi varies, with average annual precipitation around 2500mm in highland areas and 700mm in low-lying areas (Ngongondo et al. 2011).

2.1 Data sources

Data were obtained from a variety of sources and collated at the district level, as shown in Table 2.1. For our analysis, we excluded the district Likoma, an island in Lake Malawi, to give a contiguous study region. In these analyses we focused on known determinants of malaria prevalence and clinical diseases.

Table 2.1: Data sources. Climate and non-climate data variables, their description and source.

Data	Description	Spatial resolution	Temporal resolution	Source
Malaria cases	Total cases (confirmed and suspected) reported by health centres in each district	District	Monthly	HMIS
Rainfall	Rainfall estimates (mm/month)	1km grid	Monthly	CHIRPS
Min. temp	Temperature estimates (°C)	1km grid	Monthly	NOAA NCEP
Max. temp	Temperature estimates (°C)	1km grid	Monthly	NOAA NCEP
NDVI	NDVI estimates	1km grid	Monthly	LandDAAC
Population	Population estimates	District	Yearly	NSO population projections
Literacy	Proportion of population aged five and above that can read and write in any language	District	Yearly	WMS
Urban	Proportion of the population that stay in urban centres	District	Yearly	WMS
ITN	Proportion of household using ITN. The numerator is the number of sampled households that reported owning at least one ITN while the denominator is the total number of households sampled in the district.	District	Yearly	DHS
Health centres	Number of health centres that report data	District		MoH
Area	Total district area	District		Unpublished reports

2.1.1 Malaria data

We extended the previous database (Lowe et al. 2013) by adding routine malaria data for the period 2012 to 2017. We used the reported district-level monthly counts of confirmed and suspected malaria cases collected between July 2004 and December 2017, checked and cleaned by the National Malaria Control Programme (NMCP) for completeness and consistency. Case data are recorded on paper forms at a health facility within a district, then aggregated monthly at the facility level.

Facility data are subsequently aggregated to the district level and entered into an electronic database, the District Health Information System (DHIS) (*DHIS* 2018).

2.1.2 Climate data

We used satellite-derived climate archives from the library hosted at the International Research Institute (IRI). Monthly rainfall anomaly values averaged at the district level were obtained from the climate hazards group infrared precipitation with station data (CHIRPS) (*CHIRPS* 2018) which has limited station data in some countries. These data do not access extensive station data from Malawi national archives as they are not linked together. In addition, the weather station network across Malawi is sparse. Temperature anomalies were obtained from the National Oceanic and Atmospheric Administration national centres for environmental prediction (NOAA NCEP) (*NOAA* 2018). This data is based on the Climate Prediction Center (CPC) monthly global surface air temperature data set at 0.5 degrees from 1948-present. Normalised difference vegetative index (NDVI) data were collected from the LandDAAC MODIS satellite at a resolution of 1km (*LandDAAC MODIS* 2018). For the model-fitting, all gridded data were averaged over spatial areas corresponding to the districts in Malawi.

2.2 Statistical framework and model

To estimate the variation in disease risk, we modelled the *standardised morbidity ratio* (SMR). This is the ratio of observed to expected malaria cases within a single spatial unit in a single time-period and provides an estimate of the disease risk. The expected cases in each district were calculated by multiplying the district population with the annual observed risk. The annual observed risk is given by the total number of cases across all districts over the entire time period divided by the total population over the same period. SMR greater than 1 at a given time period suggests an excess risk of malaria in a district. More details on calculation of the expected cases are provided in Appendix A.

To describe the spatial and spatio-temporal variations in disease incidence we used a Poisson-log-linear mixed effects model. Let y_{st} be the observed counts in spatial unit $s = 1, \ldots, N$ and time $t = 1, \ldots, T$, and e_{st} denote the expected number of disease cases; the expected cases are calculated using standardization methods to take account of demographic differences in the populations across the different spatial units but without taking into account the effects of hypothesised risk factors or residual spatio-temporal variation (Lawson 2013). We then assumed that

$$Y_{st}|e_{st}, R_{st} \sim Poisson(e_{st}R_{st}), \tag{2.1}$$

where R_{st} is the relative risk of disease in spatial unit s at time t. In the log-linear mixed model,

$$\log(R_{st}) = x'_{st}\beta + U_{st} \tag{2.2}$$

where x_{st} is a vector of covariates (fixed effects) with associated regression parameter β and the random effects U_{st} follow a multivariate Normal distribution with zero mean vector and covariance matrix $V(\theta)$ structured to include spatial and temporal components of variation. The relative risk R_{st} is thereby decomposed into the explained and unexplained risks, $\exp(x'_{st}\beta)$ and $\exp(U_{st})$ respectively. The unexplained risk component captures residual variation after accounting for all the covariates in the model.

2.2.1 Model framework for the Malawi malaria data

The specific model formulation for the Malawi malaria data has been described in our earlier paper (Lowe et al. 2013). In brief, we first extended our notation for the model defined by 2.1 and 2.2 to distinguish between cases under and over 5 years of age. Let Y_{jst} be the monthly malaria count for age group $j = (1, 2)$ corresponding respectively to ages 5 or more and 0 to 4, district $s = 1, \ldots, m = 27$ and time $t = 1, \ldots, n = 162$ months. Similarly, let e_{jst} be the corresponding expected malaria count.

With this extended notation, we write the relative risks as $\theta_{jst} = x'_{st} + U_{st}$, where

$$U_{st} = P_s + D_t + G_{st} \tag{2.3}$$

In equation 2.3, the terms P_s, D_t and G_{st} denote purely spatial, purely temporal and residual spatio-temporal components of variation in risk, respectively. Following (Leroux et al. 2000) and (Lee et al. 2015) we assume that the G_{st} are mutually independent, $G_{st} \sim N(0, \tau_I^2)$, and that the spatial random effect, $P = (P_1, \ldots, P_m)$ and the temporal random effect, $D = D_1, \ldots, D_n$ form Gaussian Markov random fields (Rue and Held 2005). Specifically, the model defines spatial neighbourhood relationships through a symmetric $m \times m$ matrix $\mathbf{W}$ with elements $w_{ij} = 1$ if the spatial units i and j are neighbours, and $w_{ij} = 0$ otherwise; we specify i and j to be neighbours if they share a common boundary. Similarly, temporal neighbourhood relationships are defined by a symmetric $n \times n$ matrix $\mathbf{V}$; following (Lee et al. 2015) we specify $v_{ij} = 1$ if $|j - i| = 1$ and $v_{ij} = 0$ otherwise. Now, writing P_{-s} for the $(m-1)$ element vector obtained by removing the sth element from P, and similarly D_{-t} for the $(n-1)$-element vector obtained by removing the $t-th$ element from

D the model can be defined through its full conditional distributions,

$$P_s|P_{-s} \sim \mathrm{N}\left(\frac{\rho_S \sum_{j=1}^{m} w_{sj} P_j}{\rho_S \sum_{j=1}^{m} w_{sj} + 1 - \rho_S}, \frac{\tau_S^2}{\rho_S \sum_{j=1}^{m} w_{sj} + 1 - \rho_S}\right) \tag{2.4}$$

$$D_t|D_{-t} \sim \mathrm{N}\left(\frac{\rho_T \sum_{j=1}^{n} v_{tj} D_j}{\rho_T \sum_{j=1}^{n} v_{tj} + 1 - \rho_T}, \frac{\tau_T^2}{\rho_T \sum_{j=1}^{n} v_{tj} + 1 - \rho_T}\right) \tag{2.5}$$

$$\tag{2.6}$$

Both the P_s and D_t are mean-centred such that $\sum_{s=1}^{m} P_s = \sum_{t=1}^{n} D_t = 0$

We use the following diffuse prior specifications for the fixed effect parameters β and the random effect parameters $\vartheta = (\tau_S^2 \tau_T^2, \tau_I^2, \rho_S, \rho_T)$. Firstly, for the elements of β we specify independent Normal priors, $\beta_i \sim N(0, 1000) : i = 1, \ldots, p$. Secondly, for the variance components τ_S^2, τ_T^2, and τ_I^2 we specify independent inverse-Gamma priors, $\tau^2 \sim IG(1, 0.001)$. Finally, for the autocorrelation parameters ρ_S and ρ_T we specify independent uniform priors, $\rho \sim U(0, 1)$.

2.2.2 Model fitting for malaria data in Malawi

To account for differences in malaria diagnostics over time, we defined a binary variable (0 before adoption of RDTs, and 1 after adoption). Markov Chain Monte Carlo (MCMC) techniques were used to simulate from the posterior distribution using a combination of Gibbs and Metropolis-Hastings algorithms to estimate model parameters. We generated a chain of length 300,000 after a burn-in of 50,000 iterations, retaining every fiftieth iteration to obtain a sample of 5000 approximately independent realisations from the joint posterior distribution of β,θ and U for post-processing.

2.2.3 Convergence diagnostics

We carried out several tests for convergence. Trace plots for each covariate in the model were inspected. In addition, we plotted the empirical cumulative distributions of the upper and lower halves of the chains after discarding the burn-in. If the two curves overlap, it indicates that convergence was achieved. Lastly, the Geweke statistic also showed that our chains had converged. Details on the convergence diagnostics are presented in Appendix B.

3 Results

3.1 Data analyzed

Malaria case data for 27 districts analysed in this study were obtained from the HMIS for a period of 162 months ranging from July 2004 to December 2017. Prior to 2011, there was no widespread use of RDT as the policy had not been adapted. The use of RDT was adopted in 2011 leading to a marked improvement in the quality of the data. With time, the completeness of the data reported in the HMIS has been steadily going up, now standing at over 90%. Different districts have different numbers of health facilities that report data in the HMIS ranging from 4 facilities per district to over 40.

3.2 Clinical malaria patterns

The malaria case rates between the 2005 and 2017 period are shown by climatic zone in Figure 2.1. During this period the annual malaria incidence over time showed a decrease in incidence between 2009 and 2013 followed by a temporary increase from 2014 to 2015. Overall, there is a general reduction in malaria rates over the 2005-2017 period).

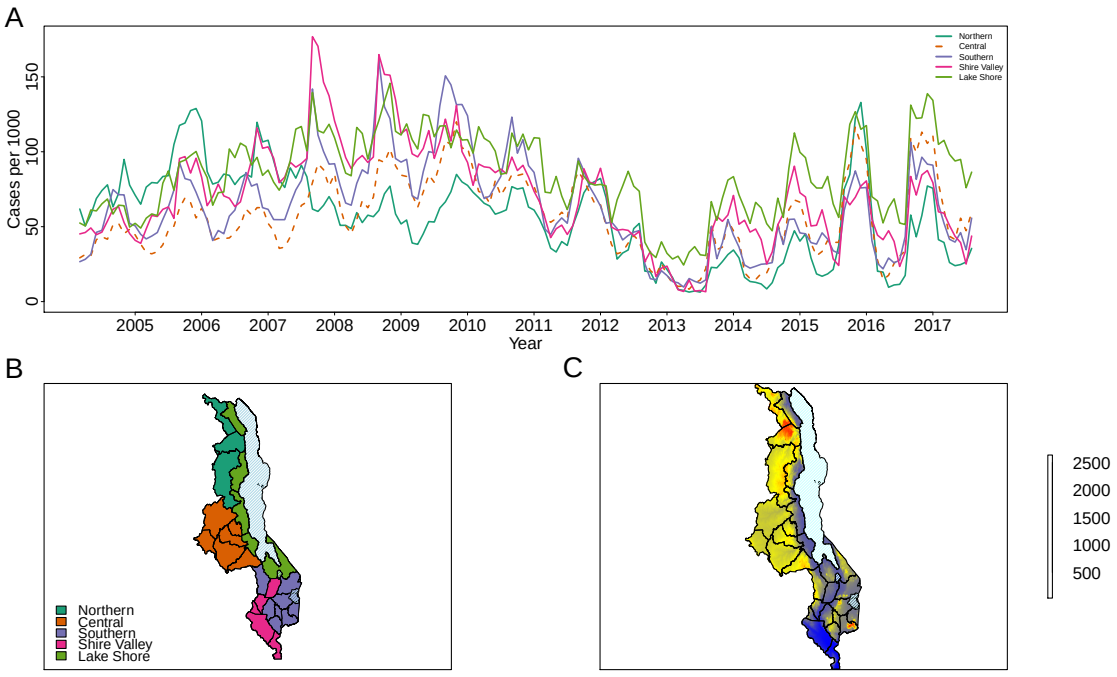

Figure 2.1: Malaria incidence patterns. (A) Mean under five monthly malaria cases for Malawi by climatic zone between July 2004 and December 2017 (B) Location of climatic zones within Malawi, (C) Altitude pattern within Malawi

Detailed seasonal patterns of malaria case rates, rainfall and temperature are

shown for each climate zone in Figure 2.2. Across zones and geographical areas, there are similar patterns of seasonality. Peak temperatures occur between October and November, before the start of the rainy season. Rainfall peaks in January, with a lag period of 0 to 3 months of peaking malaria incidence.

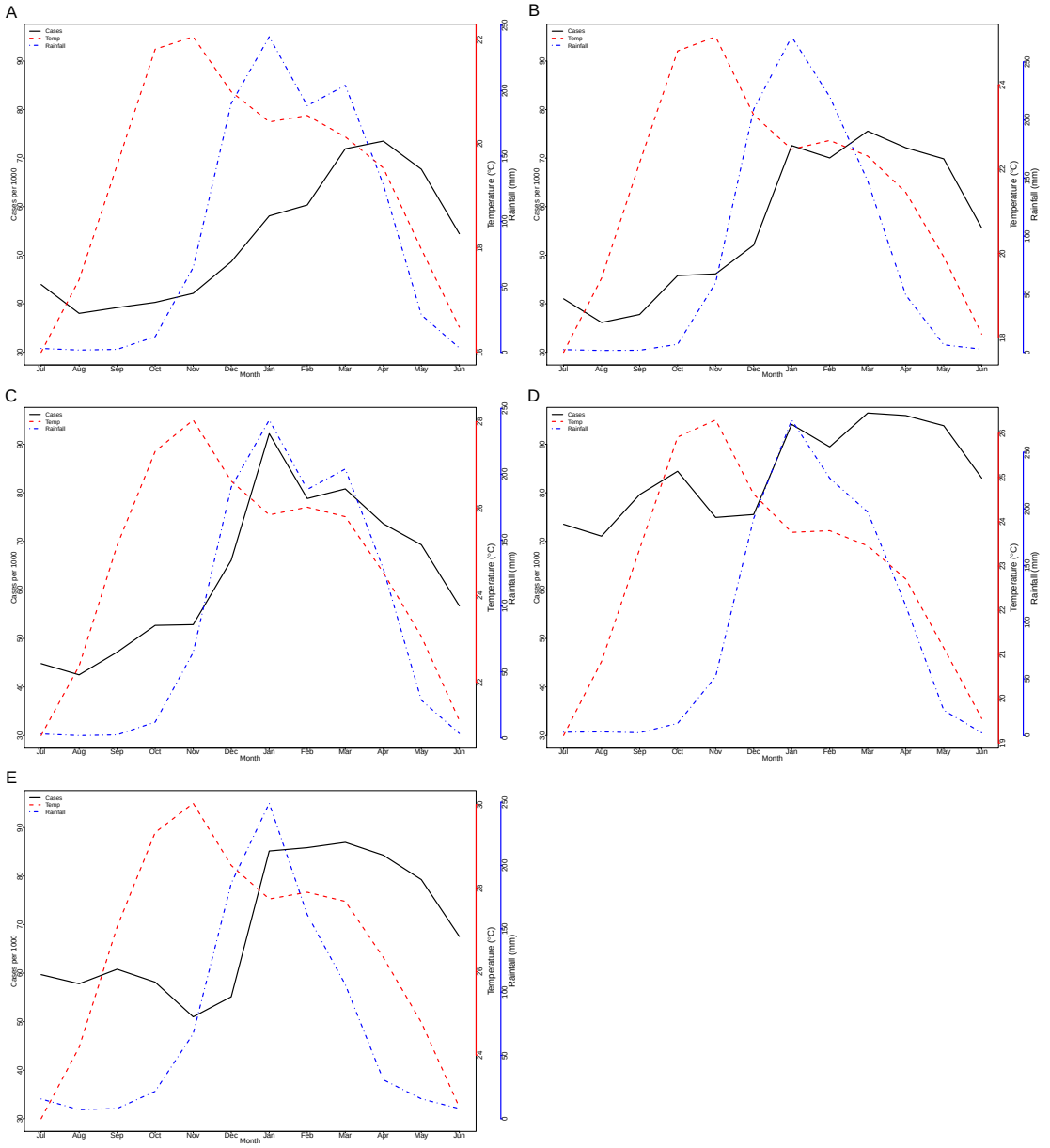

Figure 2.2: Relationship between monthly mean temperature, rainfall and malaria Monthly average malaria incidence, rainfall and temperature at the climate zonal level. (A) Northern zone (B) Central zone (C) Southern zone (D) Shire Valley (E) Lake Shore. The red dotted line is the mean temperature while the blue dotted line is the mean rainfall. The disease incidence is shown by the black solid line

Figure 2.3 shows the marginal spatial and temporal variations in malaria SMR across Malawi. The temporal variation (Figure 2.3a) indicates similar patterns of seasonality and inter-annual variation in both age groups. The spatial variation in the SMR for the age group 5 years and below (Figure 2.3b) shows higher malaria

incidence in some of the districts along the Lakeshore and Shire Valley regions. For the over-five age group (Figure 2.3c), a similar pattern is observed.

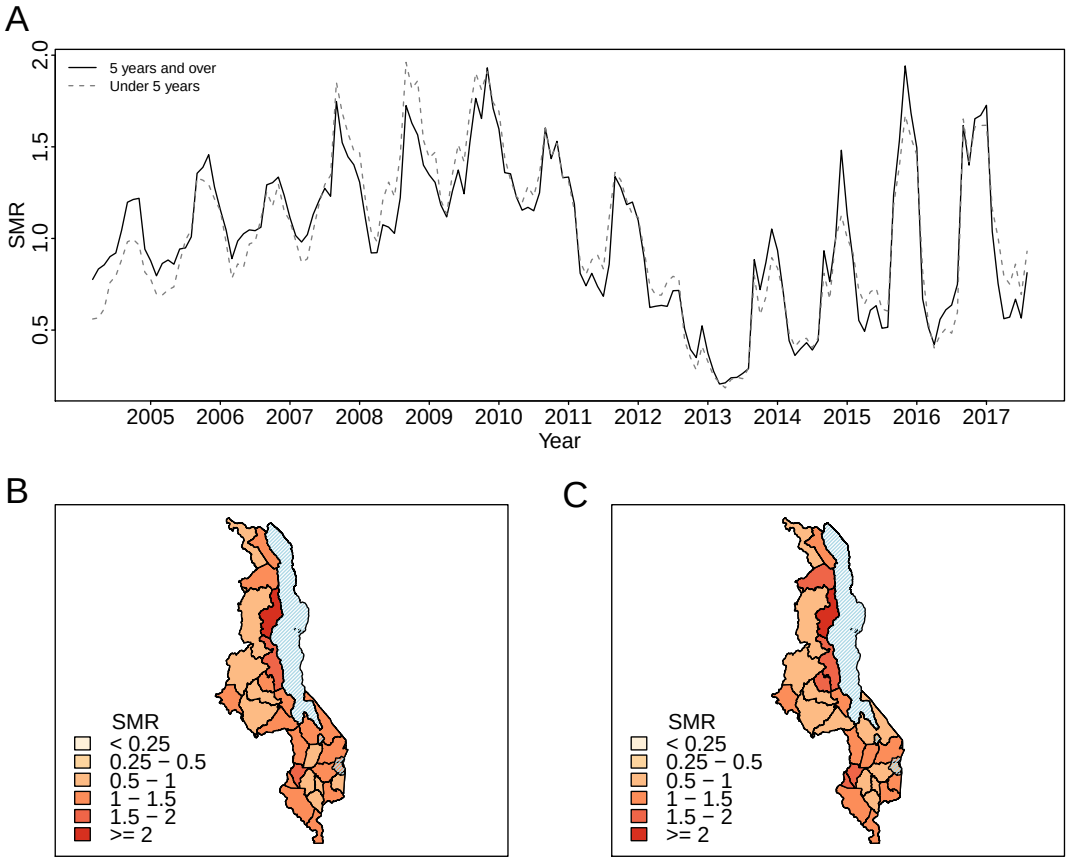

Figure 2.3: Malaria SMR averaged over time and space for the period July 2004 - December 2017 (A) averaged across the country for each month, (B) averaged over time for each district for the under 5 years age group, (C) averaged over time for each district for the age group 5 years and over

3.3 Model estimates

3.4 Association between climate and non-climate factors with clinical malaria

We first fitted a non-spatial Generalized Linear Model (GLM) to investigate the association between the outcome and different covariates and find significant predictors to include in the Generalized Linear Mixed Model (GLMM). The following covariates were included in the GLMM: mean rainfall and temperature anomalies; rainfall anomalies lagged by 1 to 3 months, temperature anomalies lagged by 1 to 3 months, NDVI, Literacy (as a proportion of the district population) and population density. We also included an indicator variable to specify the time before and after adoption of RDTs.

In the spatio-temporally structured model, clinical malaria incidence was associated with rainfall at 3 month lag, temperature including all lags and NDVI. Relative risks, 95% Bayesian credible intervals and Geweke convergence statistic G (Geweke et al. 1991) for the regression parameters are shown in Table 2.2.

Table 2.2: Parameter estimates for the mixed model. Estimates for relative risk for climatic and non-climatic parameters respectively with associated 95% credible intervals and the Geweke convergence diagnostic.

	RR	95% CI	G
Rainfall	1.00	(1.00,100)	-0.20
Rainfall lag 1	1.00	(1.00,1.00)	-0.40
Rainfall lag 2	1.00	(1.00,1.00)	-0.10
Rainfall lag 3	1.03	(1.01,1.05)	0.20
Temperature	1.03	(1.00,1.05)	-1.40
Temperature lag 1	1.03	(1.00,1.06)	0.7
Temperature lag 2	1.05	(1.03,1.08)	-1.8
Temperature lag 3	1.04	(1.01,1.07)	-1.7
NDVI	1.74	(1.45,2.07)	0.70
Literacy	1.00	(1.00, 1.00)	0.5
Pop. density	1.00	(1.00, 1.00)	-0.80
RDT	1.27	(0.96, 1.68)	0.3

After allowing for residual spatio-temporal dependence, rainfall was no longer statistically significant in the current month. However, there was a slight positive relationship between malaria incidence and rainfall in the three months prior. A unit increase in rainfall anomaly was associated with a 3% increase in malaria burden (RR=1.03, CI:1.01,1.05). For temperature anomalies in the current month, with every one-degree Celsius increase, estimated malaria incidence increased by 3% (RR=1.03, CI:1.00-1.05). Malaria was also associated with temperature anomalies at 1-3 month lags with increase in malaria of 3%, 5% and 4% respectively. NDVI was also positively associated with malaria incidence, i.e. an increase in vegetative cover is associated with a 74% increase in malaria incidence (RR=1.74, CI: 1.45-2.07). It was also observed that population density did not show an association with malaria (RR=1.00, CI: 0.99-1.00). Lastly, a 27% (RR=1.27, CI: 0.95-1.68) increase in incidence was observed in the post RDT adoption period compared to before. The Geweke diagnostic scores indicate convergence, ($-1.96 \leq G \leq 1.96$).

3.5 Mapping explained and unexplained variation in SMR

Figure 2.4 shows the decomposition of the overall risk into its explained and unexplained components. Figure 2.4a shows the overall malaria risk, R_{st} averaged over Malawi for the entire period. The final model predicts a higher than average risk in the Lakeshore and Shire Valley districts and climatic zones. In addition, 3

of the districts in the central zone also show an elevated malaria risk compared to other districts in the zone. Figure 2.4b and Figure 2.4c show the explained and unexplained component of spatial variation in risk respectively, again averaged over time. In terms of model performance, the unexplained variation $\exp(U_{st})$ is relatively high in some parts of the country indicating the presence of other district specific non-observed variables. In Figure 2.4d, it is observed that the risk explained by modelled climate covariates is almost constant across the country. In our model, we found temperature and rainfall three months prior to be significant. This shows that temperature and rainfall play a key role in malaria transmission in most of the country with a slightly different effect in the most northern and southern parts of Malawi. Lastly, non-climate covariates also contribute to the observed malaria risk roughly following the expected pattern where some of the districts along the lake and the Shire show elevated risk with notable lower risk in the major urban centres such as Blantyre.

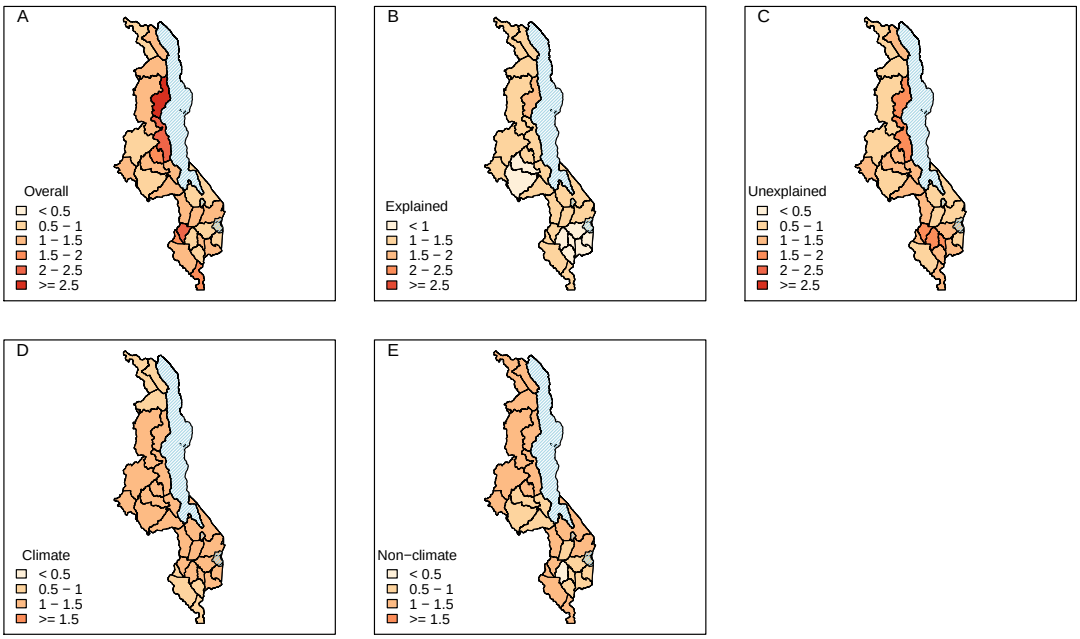

Figure 2.4: Contributions to the overall malaria risk (A) overall risk R_{st} due to combined effect of climatic, non-climatic covariates and non-observed covariates, (B) explained risk, $\exp(x'_{st}\beta)$ due to observed climatic and non-climatic covariates, (C) unexplained risk, $\exp(U_{st})$ due to unobserved effects only, (D) malaria risk due to observed climatic covariates only, (E) malaria risk due to the non-climatic covariates only

4 Discussion

Our results show the added benefit of including climate and non-climate information in modelling of malaria incidence data. Temperature and lagged rainfall were found to be significant drivers of malaria. We fitted a spatio-temporal statistical model

to quantify the effect of different climate and non-climate covariates on malaria incidence and to predict the incidence over the period July 2004 to December 2017.

While the quality and suitability of routine facility-based data is often questioned, our reported malaria incidence trends in time and space, in terms of seasonality and climate zone, largely follow the expected patterns and align with previously published malaria prevalence maps (Kazembe et al. 2006; Lowe et al. 2013). Our approach uses more recent data and covers a relatively longer time period hence similarities with the earlier maps. The Lakeshore and Shire valley areas are generally low-lying areas with higher average temperatures and higher malaria incidence.

When using routine incidence data to monitor control impact and long-term trends, control programmes need to take into account intervention implementation, climate and non-climate covariates from different data sources to improve the analysis and interpretation of disease incidence patterns.

In our study, incidence showed a steady decline in paediatric and adult data from 2009 following the introduction and gradual scale up of efficacious artemisinin-based combination therapy (ACT) treatment in 2008 and start of ITN distribution to mothers and children. This decline was followed by a nadir and upsurge after 2013. While similar reversals and increases have been reported by countries in the southern African region (Nkumama et al. 2017), their control programmes included different control implementation stages during this period, suggesting other factors could be at play. Analysing climate and non-climate factors, and visualizing the explained and unexplained components of the observed variation in disease risk, can give additional insight. While the introduction and scale up of malaria rapid diagnostic tests (mRDTs) and inclusion of community-based malaria treatment from 2012 could have led to increased health case seeking behaviour and capture of cases that previously did not present to the health care system, the period after 2013 also documented higher average temperatures.

In terms of model performance, the unexplained variation is lower across the country. This shows that most of the variation has been captured by the covariates in the model. However, the substantial unexplained risk shows the importance of including random effects in the model.

The non-significance of rainfall on malaria in the current month in our analyses shows the complexity of the relationship between malaria incidence and climate in general, but rainfall in particular. Studies in different settings have shown mixed effects of rainfall on malaria; some have shown a positive association, whilst others have shown a very weak or no association (Hoek et al. 1997; Lindsay et al. 2000; Abeku et al. 2003). NDVI is also significantly associated with malaria. Seasonal and year to year changes in NDVI are commonly associated with rainfall. Green vegetative cover, which is prevalent in the rainy season, is positively associated

with malaria incidence. Several other studies have shown vegetative cover to be a significant predictor (Gaudart et al. 2009; Fastring and Griffith 2009). Temperature plays key role in the development of malaria vectors and their activities that directly or indirectly lead to the spread of malaria. It has been found to be a significant ecological factor in several studies such as (Paaijmans et al. 2009; Beck-Johnson et al. 2017; Stresman 2010) but its impact on malaria transmission in tropical climates is usually considered a highland phenomenon (Lyon et al. 2017).

Both components of the decomposed risk (overall and covariate-explained) could be affected by other important factors that were not considered in these analyses. This include the completeness and quality control of monthly reports from government and faith-based health facilities that do report into the DHIS2 system, but could also come from other health facilities, mainly private for-profit, that do not report their data to the MoH via the DHIS2. While we did not have access to data on reporting completeness or quality at district level, it is likely that reporting rates influence the estimation of the malaria burden in Malawi.

While our analyses show how climate and non-climate data from multiple sources can be used to improve the analysis and interpretation of routine malaria data patterns, we are clear on the limitations and strengths of the Malawi data over the reporting period and potential steps moving forward.

Self-treatment at home will never be captured in the HMIS. Any substantial changes in the proportion of home-treatment within the country over time could affect routine facility-based disease trends. In Malawi, however, the availability of antimalarials in rural areas is limited and treatment is provided for free by the government. The introduction and scale up of RDTs in 2011 and the programmes and steps to link the reported diagnosis and treatment to consumables stock management over the past year, provide reassurance on the reported cases moving forward. Prior to 2011, when the MoH adopted the policy of testing all suspected cases by RDT, (Ministry of Health (MOH) 2011), presumptive treatment of malaria was widespread in Malawi. This may have affected the accuracy of the reported cases in the period before 2011.

Selection bias in seeking health care due to differential access to health facilities among different groups of people and variable distances between facilities and homes is another common concern with routine facility-based data (Amouzou et al. 2013). People living very far away from health facilities from may not be adequately represented in routine data. Actually, the inclusion of community-level treatment of malaria cases based on mRDTs by community health workers in hard-to-reach areas in Malawi has been included in reporting to the DHIS2 since 2012 and may have contributed to the increase in reported cases across the country, but as they are included within the health facility level reports for the relevant catchment area,

it was not possible to confirm this in the current analyses.

We used satellite-derived climate data for our models, rather than directly measured climate data from weather stations. Ideally, a high-quality gridded climate database including rainfall and temperature (minimum and maximum) from weather stations should be used to formulate models and produce malaria risk predictions. National climate data sets which integrate global products and all relevant local observations managed by the national meteorological agencies are increasingly available in African countries (Dinku et al. 2016).

Intervention coverage status data was not available at district level for the period of interest, as this data is not part of the routine data collection and is assessed at regional level in the national household malaria indicator surveys. We relied on crude intervention implementation proxies in the presented model.

Despite these limitations, the presented work shows the potential added value of our spatio-temporal statistical modelling approach. Furthermore, there are three promising developments in Malawi that will soon offer opportunities to apply the framework with more detailed data on key covariates. First, as part of a collaboration with the LINK programme in Malawi (*LINK* 2018) intervention coverage maps will soon become available for key interventions including ACTs, mRDTs and ITNs, allowing integrating coverage scale-up. Secondly, as the LINK programme modelled spatio-temporal prevalence data at district level, we will have the opportunity for more comparative analyses of modelled transmission and burden data. Lastly, electronic facility level reporting of clinical cases into the DHIS2 began in 2018, which will soon allow more granular mapping of disease risk at health facility catchment area, providing the opportunity to analyse more detailed spatial patterns moving forward. With these developments, the presented model framework can be expanded towards more in-depth analyses of intervention impact.

5 Conclusion

This work provides a modelling framework for integrating climatic and non-climatic information into analyses of routine malaria case data at facility-level, in order to improve understanding of climate effects on climate-sensitive VBD such as malaria, while simultaneously controlling for non-climatic risk factors. The findings show the value of collaborations between control programmes, health researchers and climate experts in the collation, analyses and interpretation of routine malaria data. Visualizing the findings in maps produced provide an easy to use tools for malaria control programmes to support their interpretation of disease trends over time,

which, with the development of user friendly analysis tools could be incorporated into Technical Working Groups (TWGs) and standard programme review processes.

References

Abeku, Tarekegn Abose, Gerrit J van Oortmarssen, Gerard Borsboom, Sake J de Vlas, and JDF Habbema (2003). "Spatial and temporal variations of malaria epidemic risk in Ethiopia: factors involved and implications." In: *Acta tropica* 87.3, pp. 331–340.

Altizer, Sonia, Richard S Ostfeld, Pieter TJ Johnson, Susan Kutz, and C Drew Harvell (2013). "Climate change and infectious diseases: from evidence to a predictive framework." In: *science* 341.6145, pp. 514–519.

Amouzou, Agbessi, Willie Kachaka, Benjamin Banda, Martina Chimzimu, Kenneth Hill, and Jennifer Bryce (2013). "Monitoring child survival in 'real-time' using routine health facility records: results from Malawi." In: *Tropical Medicine & International Health* 18.10, pp. 1231–1239.

Beck-Johnson, Lindsay M, William A Nelson, Krijn P Paaijmans, Andrew F Read, Matthew B Thomas, and Ottar N Bjørnstad (2017). "The importance of temperature fluctuations in understanding mosquito population dynamics and malaria risk." In: *Open Science* 4.3, p. 160969.

Bødker, Ren, William Kisinza, Robert Malima, Hamisi Msangeni, and Steve Lindsay (2000). "Resurgence of malaria in the Usambara mountains, Tanzania, an epidemic of drug-resistant parasites." In: *Global Change and Human Health* 1.2, pp. 134–153.

Cash, BA, X Rodó, J Ballester, MJ Bouma, A Baeza, R Dhiman, and M Pascual (2013). "Malaria epidemics and the influence of the tropical South Atlantic on the Indian monsoon." In: *Nature climate change* 3.5, pp. 502–507.

CHIRPS (2018). URL: http://iridl.ldeo.columbia.edu/SOURCES/.UCSB/.CHIRPS/.v2p0/.monthly/.global/.precipitation (visited on 09/03/2018).

Connor, Stephen J and Gilma C Mantilla (2008). "Integration of seasonal forecasts into early warning systems for climate-sensitive diseases such as malaria and dengue." In: *Seasonal Forecasts, Climatic Change and Human Health*. Springer, pp. 71–84.

DHIS (2018). URL: dhis2.health.gov.mw/ (visited on 09/03/2018).

Diggle, Peter, Rana Moyeed, Barry Rowlingson, and Madeleine Thomson (2002). "Childhood malaria in the Gambia: a case-study in model-based geostatistics." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51.4, pp. 493–506.

Dinku, T, R Cousin, J Corral, and A Vadillo (2016). "The ENACTS Approach: Transforming climate services in Africa one country at a time." In: *World Policy Papers*, pp. 1–24.

Fastring, DR and JA Griffith (2009). "Malaria incidence in Nairobi, Kenya and dekadal trends in NDVI and climatic variables." In: *Geocarto International* 24.3, pp. 207–221.

Gaudart, Jean, Ousmane Touré, Nadine Dessay, A lassane Dicko, Stéphane Ranque, Loic Forest, Jacques Demongeot, and Ogobara K Doumbo (2009). "Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali." In: *Malaria Journal* 8.1, p. 61.

Geweke, John et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Vol. 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.

Githeko, Andrew K, Steve W Lindsay, Ulisses E Confalonieri, and Jonathan A Patz (2000). "Climate change and vector-borne diseases: a regional analysis." In: *Bulletin of the World Health Organization* 78.9, pp. 1136–1147.

Hoek, W Van Der, F Konradsen, D Perera, Priyanie H Amerasinghe, and Felix P Amerasinghe (1997). "Correlation between rainfall and malaria in the dry zone of Sri Lanka." In: *Annals of Tropical Medicine & Parasitology* 91.8, pp. 945–949.

Kazembe, Lawrence N, Immo Kleinschmidt, Timothy H Holtz, and Brian L Sharp (2006). "Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data." In: *International Journal of Health Geographics* 5.1, p. 41.

Lafferty, Kevin D (2009). "The ecology of climate change and infectious diseases." In: *Ecology* 90.4, pp. 888–900.

LandDAAC MODIS (2018). URL: http://iridl.ldeo.columbia.edu/SOURCES/.USGS/.LandDAAC/.MODIS/.version_005/.SAF/ (visited on 09/03/2018).

Lawson, Andrew B (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.

Lee, Duncan, Alastair Rushworth, and Gary Napier (2015). "CARBayesST: An R Package for Spatio-temporal Areal Unit Modelling with Conditional Autoregressive Priors." In:

Leroux, Brian G, Xingye Lei, and Norman Breslow (2000). "Estimation of disease rates in small areas: a new mixed model for spatial dependence." In: *Statistical models in epidemiology, the environment, and clinical trials*. Springer, pp. 179–191.

Lindblade, Kim A, Edward D Walker, Ambrose W Onapa, Justus Katungu, and Mark L Wilson (2000). "Land use change alters malaria transmission parameters

by modifying temperature in a highland area of Uganda." In: *Tropical Medicine & International Health* 5.4, pp. 263–274.

Lindsay, Steven W, René Bødker, Robert Malima, Hamisi A Msangeni, and Wiliam Kisinza (2000). "Effect of 1997–98 EI Niño on highland malaria in Tanzania." In: *The Lancet* 355.9208, pp. 989–990.

LINK (2018). URL: https://www.linkmalaria.org/ (visited on 09/03/2018).

Lowe, Rachel, Trevor C Bailey, David B Stephenson, Richard J Graham, Caio AS Coelho, Marilia Sa Carvalho, and Christovam Barcellos (2011). "Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil." In: *Computers & Geosciences* 37.3, pp. 371–381.

Lowe, Rachel, Christovam Barcellos, Caio AS Coelho, Trevor C Bailey, Giovanini Evelim Coelho, Richard Graham, Tim Jupp, Walter Massa Ramalho, Marilia Sá Carvalho, David B Stephenson, et al. (2014). "Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts." In: *The Lancet Infectious Diseases* 14.7, pp. 619–626.

Lowe, Rachel, James Chirombo, and Adrian M Tompkins (2013). "Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi." In: *Malaria journal* 12.1, p. 1.

Lyon, Bradfield, Tufa Dinku, Anita Raman, and Madeleine C Thomson (2017). "Temperature suitability for malaria climbing the Ethiopian Highlands." In: *Environmental Research Letters* 12.6.

Ministry of Health (MOH) (2011). "Guidelines for use of malaria rapid diagnostic tests (mRDTs) in Malawi." In:

Ngongondo, Cosmo, Chong-Yu Xu, Lars Gottschalk, and Berhanu Alemaw (2011). "Evaluation of spatial and temporal characteristics of rainfall in Malawi: a case of data scarce region." In: *Theoretical and applied climatology* 106.1-2, pp. 79–93.

Nkegbe, Paul Kwame, Naasegnibe Kuunibe, and Samuel Sekyi (2017). "Poverty and malaria morbidity in the Jirapa District of Ghana: A count regression approach." In: *Cogent Economics & Finance* 5.1, p. 1293472.

Nkumama, Irene N, Wendy P O'Meara, and Faith HA Osier (2017). "Changes in malaria epidemiology in Africa and new challenges for elimination." In: *Trends in parasitology* 33.2, pp. 128–140.

NMCP, National Malaria Control Programme (2011). *Malawi Malaria Indicator Survey 2010*. Lilongwe,Malawi: NMCP.

NMCP, National Malaria Control Programme and ICF (2018). "Malawi Malaria Indicator Survey 2017." In:

NOAA (2018). URL: https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.GHCN_CAMS/.gridded/.deg0p5 (visited on 09/03/2018).

Paaijmans, Krijn P, Andrew F Read, and Matthew B Thomas (2009). "Understanding the link between malaria risk and climate." In: *Proceedings of the National Academy of Sciences* 106.33, pp. 13844–13849.

Paaijmans, Krijn P, Moses O Wandago, Andrew K Githeko, and Willem Takken (2007). "Unexpected high losses of Anopheles gambiae larvae due to rainfall." In: *PLoS One* 2.11, e1146.

Parham, Paul Edward and Edwin Michael (2010). "Modeling the effects of weather and climate change on malaria transmission." In: *Environmental health perspectives* 118.5, p. 620.

Rue, Havard and Leonhard Held (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

Snow, Robert W, Benn Sartorius, David Kyalo, Joseph Maina, Punam Amratia, Clara W Mundia, Philip Bejon, and Abdisalan M Noor (2017). "The prevalence of Plasmodium falciparum in sub-Saharan Africa since 1900." In: *Nature* 550.7677, p. 515.

Stewart-Ibarra, Anna M and Rachel Lowe (2013). "Climate and non-climate drivers of dengue epidemics in southern coastal Ecuador." In: *The American journal of tropical medicine and hygiene* 88.5, pp. 971–981.

Stresman, Gillian H (2010). "Beyond temperature and precipitation: ecological risk factors that modify malaria transmission." In: *Acta Tropica* 116.3, pp. 167–172.

Thomson, Madeleine C, Simon J Mason, Thandie Phindela, and Stephen J Connor (2005). "Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana." In: *The American journal of tropical medicine and hygiene* 73.1, pp. 214–221.

Thomson, MC, FJ Doblas-Reyes, SJ Mason, R Hagedorn, SJ Connor, T Phindela, AP Morse, and TN Palmer (2006). "Malaria early warnings based on seasonal climate forecasts from multi-model ensembles." In: *Nature* 439.7076, pp. 576–579.

Tompkins, Adrian M, Laragh Larsen, Nicky McCreesh, and David Taylor (2016). "To what extent does climate explain variations in reported malaria cases in early 20th century Uganda?" In: *Geospatial health* 11.1s.

Townes, Lindsay R, Dyson Mwandama, Don P Mathanga, and Mark L Wilson (2013). "Elevated dry-season malaria prevalence associated with fine-scale spatial patterns of environmental risk: a case–control study of children in rural Malawi." In: *Malaria journal* 12.1, p. 1.

Wu, Xiaoxu, Yongmei Lu, Sen Zhou, Lifan Chen, and Bing Xu (2016). "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation." In: *Environment international* 86, pp. 14–23.

Yadav, Kavita, Sunil Dhiman, Bipul Rabha, PK Saikia, and Vijay Veer (2014). "Socio-economic determinants for malaria transmission risk in an endemic primary health centre in Assam, India." In: *Infectious diseases of poverty* 3.1, p. 19.

Chapter 3

Estimation of spatially varying effective reproduction numbers for infectious disease epidemics

Chirombo J, Diggle P.J., Terlouw D.J., Read J.M.

3. Estimation of spatially varying effective reproduction numbers for infectious disease epidemics

Chirombo J, Diggle P.J., Terlouw D.J., Read J.M.

Abstract

The effective reproduction number, $R(t)$, is defined as the average number of secondary infections caused by a newly infected individual at time t. It is an important epidemiological measure of the growth of an infectious disease epidemic as it can be used to monitor and characterize outbreaks, and to provide near-real-time assessment of the efficacy of control interventions. When $R(t) < 1$ the epidemic can be said to be under control. Most methods developed to estimate reproduction numbers ignore geographical patterning of cases. The ability to characterize the transmission potential within and between different sub-regions that together make up the region of interest would be helpful to identify sub-regions where $R(t) > 1$ and to prioritize control efforts during an outbreak. We present an extension of established methods to estimate the reproduction number in a way that incorporates spatial as well as temporal information on cases to derive estimates of the epidemic growth across geographical space. To illustrate the methodology, we apply the new method to case data from the 2014-15 West African Ebola epidemic.

1 Introduction

Quantifying the transmissibility of infectious diseases as they progress is paramount in the planning and application of control measures. The effective reproduction number $R(t)$, defined as the average number of secondary infections caused by a newly infected individual at time t, gives a measure of the overall state of the epidemic, with $R(t) > 1$ indicating a growing epidemic and $R(t) < 1$ an epidemic that is under control. Changes in $R(t)$ reflect the changing dynamics of the epidemic due to control measures, depletion of susceptible individuals in the population and other factors including response of individuals to the epidemic, irrespective of control effort. For instance, in the 2014-2015 West Africa Ebola epidemic it is likely that some communities reduced their contact rates at particular times in response to the state of the epidemic in neighbouring communities, while control remained constant over the same time period.

Several methods have been developed to calculate $R(t)$. For example, Wallinga and Teunis (2004) developed a likelihood-based method that infers the reproduction number from the observed epidemic curves. By considering pairs of cases, they were able to calculate the relative likelihoods of different infected individuals transmitting the infection to a susceptible individual. This method, however, does not deal with the problem of censoring, whereby observations beyond t are needed to calculate the reproduction number at t. It also describes $R(t)$ by considering cohorts of cases whose symptom onset is on the same day, and therefore cannot say anything about $R(t)$ on days where there are no new cases. Cauchemez et al. (2006) have developed a Bayesian method for estimating $R(t)$ in real-time from the observed epidemic curve together with partial tracing information. Fraser (2007) developed a method for calculating both the cohort and instantaneous reproduction numbers. The cohort reproductive number estimates how many people each case actually infects and can only be applied retrospectively while instantaneous reproductive number is the average number of secondary cases that each infected individual would infect if the conditions remained as they were at time t. Cori et al. (2013) extended the work of Fraser (2007) to develop a method for estimating an instantaneous $R(t)$ in near-real-time allowing for the effect of censoring.

In all the methods described above, the estimate of $R(t)$ is an average over the entire geographical region of interest, ignoring any heterogeneity in disease transmission potential across geographical space. Possible causes of such heterogeneity include spatial variation in demographic, environmental and socio-economic factors. Estimates that takes into account spatial heterogeneity would help authorities to prioritize the application of interventions, especially in resource-poor settings. White et al. (2013) incorporated spatial information in the estimation of $R(t)$ for influenza in South Africa. They modified the method proposed by Wallinga and Teunis (2004) by defining a spatial metric based on distances and travel patterns between the South African provinces and allowing the transmission probabilities between provinces to depend on this metric. However, their method does not account for the censoring of future cases.

In this paper we consider the geographical region of interest to be partitioned into m sub-regions, labelled $x_1, \ldots, x_m$ and develop an estimator for a spatio-temporal reproduction number $R(x, t)$, defined as the total number of secondary infections caused by an individual in sub-region x who becomes infectious at time t, allowing for the censoring of future cases.

The structure of the paper is as follows. In Section 2, we set out the definitions that we need in later sections. In Section 3, we review the method proposed by Wallinga and Teunis (2004) for calculating $R(t)$. In Section 4, we show how this method can be extended to incorporate spatial information and define a class of

parametric models for the transmission from a primary case in one sub-region to a secondary case in another region. In Section 5 we show how to fit the parametric model whilst allowing for the censoring of future cases. In Section 6 we apply our method to calculate reproduction numbers for the 2014-15 Ebola epidemic in West Africa. The paper ends with a discussion in Section 7.

2 Definitions

We write $n(x, T]$ for the number of cases that occur within a geographical region x and time-interval $[0, T]$. Each case can be associated with a number of distinct event-times, including: time of infection; time of symptom onset; time of reporting; beginning and end of infectious period. In practice, not all of these quantities are observable. Here, we assume that each case is associated with a single, observed *incident time* $t_i : i = 1, ..., n$, and that individual cases are ordered according to their incident times. We also assume that each case has an associated symptom onset time $s_i : i = 1, \ldots, n$. In practice, times are always recorded in discrete units, for example days; for the time being, we assume that coincident cases are ordered arbitrarily.

The *infector* of case i is denoted by $v(i)$. The ith *generation interval*, U_i, is the time-lag between t_i and its infector, hence $U_i = t_i - t_{v(i)} : i = 2, 3, ..., n$. This is equal to $t_i - t_j$ for some unknown value of $j < i$. Similarly, the ith *serial interval*, D_i, is the time lag between s_i and its infector, i.e $D_i = s_i - s_{v(i)} : i = 2, 3, \ldots, n$. The U_i and D_i are therefore random variables, which we assume to be independently and identically distributed according to a specified probability density functions $f(u) : u \geq 0$ and $v(d) : d \geq 0$. In practice however, the actual infection times, t_i are not observed but rather symptom onset times s_i. Therefore, the serial interval distribution is widely used in place of the generation interval and they are used interchangeably. Depending on the substantive meaning of each t_i, $f(u)$ is known as the *generation interval distribution* or the *serial interval distribution* (hereafter SI distribution). The *epidemic curve* is the cumulative number of incident cases at or before time t.

3 The Wallinga-Teunis method

The Wallinga-Teunis method (hereafter WT) estimates $R(t)$ using a probability-based argument. The method uses the epidemic curve and the SI distribution. It assumes that recorded incident times are discrete. This is almost always the case in practice; if not, it can be achieved by grouping the incident times into a discrete

set of intervals. Given the recorded incident times, and assuming perfect mixing within the population at risk, the probability p_{ij} that j is the infector of i is

$$p_{ij} = \frac{f(t_i - t_j)}{\sum_{r<i} f(t_i - t_r)} : j = 1, ..., i-1 \tag{3.1}$$

An estimate of the total number of cases infected by case j is

$$R_{0j} = \sum_{i>j} p_{ij}. \tag{3.2}$$

The corresponding estimate of the reproduction number at time t is

$$R(t) = m_t^{-1} \sum_{j=1}^{m_t} R_{0j}, \tag{3.3}$$

the sample mean of the R_{0j} over the m_t individuals j for whom $t_j = t$. This is called the *cohort* reproduction number as it estimates the average number of infections per member of the cohort of cases incident at time t (Cori et al. 2013).

4 A spatial extension of the Wallinga method

Note, firstly, that the WT method can be derived from a simple application of Bayes' Theorem as follows. The conditional probability density of the incident time t_i, given that j, occurring at time $t_j < t_i$, is the infector of i, is $f(t_i - t_j)$. Then, Bayes' Theorem states that

$$\begin{aligned} \mathrm{P}(j \text{ is infector of } i | t_i) &= \frac{f(t_i - t_j)\mathrm{P}(j \text{ is infector of } i)}{\sum_{k<i} f(t_i - t_k)\mathrm{P}(k \text{ is infector of } i)} \\ &= \frac{f(t_i - t_j)}{\sum_{k<i} f(t_i - t_k)} \end{aligned} \tag{3.4}$$

because, under perfect mixing, all preceding cases are equally likely *a priori* to be the infector of case i .

Now, using x_i to denote the sub-region in which case i occurs and writing q_{ij} as a short-hand for the conditional probability that j is the infector of i given the incident times and locations up to and including case i, the spatial counterpart of (3.4) is

$$q_{ij} = \frac{f(t_i - t_j)p(x_i, x_j)}{\sum_{k<i} f(t_i - t_k)p(x_i, x_k)}, \tag{3.5}$$

where $p(x_i, x_j)$ is a *transmission probability*, the *a priori* probability that the infector of a case at location x_i will be from x_j. This formulation assumes perfect mixing and the same SI distribution *within* each location. It reduces, as it must, to WT

when there is only one location.

Now, writing $R_{0j}(x)$ for the expected number of cases in sub-region x infected by case j, we have that

$$R_{0j}(x) = \sum_{i>j} q_{ij}. \tag{3.6}$$

The estimated average number of subsequent cases at location x per incident case at time t and location y, are then the sample means of the $R_{0j}(x)$ over the m_{yt} individuals j at location y whose incident time is t, hence

$$R(x, y, t) = m_{yt}^{-1} \sum_j R_{0j}(x) I(x_j = y) I(t_j = t), \tag{3.7}$$

where $I(\cdot)$ is the indicator function. The *spatial reproduction number*, $R(y, t)$, defined as the expected number of future cases generated by an incident case at location y and time t follows from (3.7) by summing over all locations x, hence

$$R(y, t) = \sum_x R(x, y, t) \tag{3.8}$$

In principle, we could implement (3.8) by specifying any mathematically valid SI distribution $f(u)$ and transmission probabilities $p(x, y)$. For this to be a useful exercise, we need to estimate these quantities, either using data from previous studies together with scientific knowledge of the disease in question, or directly from the current data. In Section 5 we consider the possibilities and limitations of the latter approach.

5 Parametric model formulation

The quantities so far defined, namely the SI distribution $f(u)$ and the transmission probabilities $p(x, y)$, do not define a complete model for an epidemic process; rather, they are properties of the epidemic process. However, these properties do determine the quantities that are of epidemiological interest here, namely the spatial reproduction numbers $R(y, t)$.

5.1 Modelling the transmission probabilities

Our proposed model assumes that the rate at which a susceptible individual in location x gets an infection from an infectious individual in location y at time t factorises as $\lambda(x, y, t) = \lambda_0(t)\nu(x, y)$ where

$$\nu(x, y) = \exp(A_x + B_y + C_{xy}), \tag{3.9}$$

and therefore;

$$\lambda(x, y, t) = \lambda_0(t) \exp(A_x + B_y + C_{xy}) \tag{3.10}$$

In equation 3.10, the terms A_x, B_y and C_{xy} represent the susceptibility of individuals at location x, the infectiousness of individuals at location y and the connectivity between locations x and y, and $\lambda_0(t)$ is the baseline hazard of infection at time t, which we assume to be common to all locations.

The total rate of transmission from y to x is proportional to $N_x N_y \nu(x, y)$, where N_x is the number of susceptibles at location x and N_y is the number of infected at location y. The corresponding transmission probabilities follow as

$$p(x, y) = N_y \nu(x, y) / \sum_z N_z \nu(x, z) \tag{3.11}$$

It follows from (3.5) and (3.11) that if we know the population sizes N_x and can estimate ratios of the quantities $\nu(x, y)$, we can estimate the probabilities q_{ij}, and the spatial reproduction numbers follow by substitution into (3.6), (3.7) and (3.8).

One possible parametric version of (3.9) is a log-linear formulation,

$$\nu(x, y) = \exp(a'_x \alpha + b'_y \beta + c'_{xy} \gamma) \tag{3.12}$$

where now a_x, b_y and c_{xy} are vectors of explanatory variables with associated parameters α, β and γ respectively. Given sufficient data, this model could be further extended to allow individual-level explanatory variables, e.g. age or gender, but we do not pursue this here.

The *gravity model* is the special case of (3.12) in which $c_{xy} = \log d(x, y)$ where $d(\cdot)$ denotes distance. In the absence of covariates, this equates to

$$\nu(x, y) = \exp(\alpha + \beta) / d(x, y)^\gamma. \tag{3.13}$$

Other choices for term $h(x, y) = c'_{xy}\gamma$ in (3.9) can be obtained as special cases of a general transmission kernel model, for which

$$\nu(x, y) = \left(\delta + \frac{d(x, y)}{\sigma} \right)^{-\gamma}$$

For example, $\delta = 0$ and $\sigma = 1$ gives the *power law*, whilst $\alpha \neq 0$ and $\sigma \neq 0$ gives a *lagged power law* model Meyer, Held, et al. (2014).

5.2 Modelling the serial interval distribution

Parametric models for the SI such as the gamma, weibull and lognormal distributions are widely used to model infectious diseases such as influenza (Cowling et al. 2009;

Cauchemez et al. 2006). It is difficult to estimate both the SI distribution and the transmission kernel from one dataset. Better results can therefore be expected if it is possible to use either contextual knowledge or other data to specify one of the two.

Here, we assume a gamma distribution for the SI distribution. This two-parameter family can generate distributions with a range of positive skewness and includes the symmetric Normal distribution as a limiting case. Other possible families of distributions include the Weibull and log-Normal. These are also two-parameter families that generate positively skewed distributions, and are hard to distinguish empirically from the gamma distribution (Firth 1988).

5.3 Parameter estimation and model selection

To estimate the parameters of the model, θ say, from current data we use an adaptation of the partial likelihood method introduced by Cox (1972); see also Diggle et al. (2010). Consider the sequence of incident case times t_i and associated sub-regions $x_i : i = 1, ..., s$. Write N_i for the number of individuals at risk in sub-region x_i, and $I_i(t)$ for the number of infectious individuals in sub-region x_i at time t. Recall that we can impose a unique ordering of the cases according to their incident times by randomly permuting any sets of cases with identical t_i. The partial likelihood for θ is the likelihood of the observed ordering of the x_i, which we derive as follows. The rate at which any sub-region x_i acquires a new case at time t_i is $\rho_i(t) = \sum_{j=1}^{s} \nu(x_i, x_j) N_i I_j(t_i)$. The probability that the ith time-ordered case occurs in sub-region x_i is

$$P_i = \left(\rho_i(t_i) / \sum_{k=1}^{s} \rho_k(t_i) \right) \tag{3.14}$$

and the partial log likelihood is given by the expression;

$$PL(\theta) = \sum_{i=1}^{n} \log \left(\rho_i(t_i) / \sum_{k=1}^{s} \rho_k(t_i) \right) \tag{3.15}$$

It follows from (3.15) that the partial likelihood can only identify a sub-set of the model parameters θ, namely those that are not cancelled when taking ratios of the $\nu(x, y)$. But in the current context this does not matter, because the ratios are all we need to estimate. For example, under model (3.13)

$$PL(\theta) = \sum_{i=1}^{n} \log \left(\frac{\sum_{j=1}^{s} d(x_i, x_j)^{-\gamma} N_i I_j(t_i)}{\sum_{k=1}^{s} \sum_{j=1}^{s} d(x_k, x_j)^{-\gamma} N_k I_j(t_i)} \right)$$

Partial likelihood is beneficial over full maximum likelihood estimation in several

ways. Firstly, it is easier to implement especially in situations where there are nuisance parameters which are unidentifiable. This property is attractive when these parameters are not of interest hence reducing the computation burden. Partial likelihood has also been found to perform better with smaller sample sizes compared to the full maximum likelihood estimation which requires relatively larger sample sizes to achieve less bias (Lin and Zhu 2012). Partial likelihood estimates have the same asymptotic properties as maximum likelihood estimators. However, there may be loss of efficiency and some parameters of interest may be unidentifiable (Diggle 2006).

6 Application: the 2013-16 West Africa Ebola epidemic

Ebola virus disease (EVD) is a highly infectious disease with a high fatality rate Shultz et al. (2016). The West African Ebola outbreak of 2013-2016 was the biggest Ebola outbreak in history and the first to attack West Africa. The disease was first reported in Guinea in December 2013 (WHO 2014). The World Health Organisation (WHO) declared the epidemic to be a public health emergency in August 2014. Liberia, Guinea and Sierra Leone were the worst affected countries in the region. Between December 2013 and April 10, 2016, a total of 28,616 cases were reported with 11,310 deaths (WHO 2016). The peak of the epidemic was in September 2014 when 950 cases were reported in one week (WHO 2016). The three affected countries in West Africa all share land boundaries and therefore there was potential for cross-border spread of the disease.

6.1 Exploratory analysis

The EVD had a mean incubation period of 11.4 days in each of the 3 countries. Approximately 95% of patients experienced symptom onset within 21 days after exposure (WHO 2014). The SI, estimated from a subset of the cases whose information on symptom onset dates or suspected transmission chains were available, was estimated to be 15.3 days with a standard deviation of 9.3 days (WHO 2014). The longer SI compared to past outbreaks and the highly dispersed SI distribution possibly point to difficulties in collecting unbiased data on exposure through contact tracing during the 2013-16 outbreak or past outbreaks. It may also indicate a higher proportion of transmission events occurring late during an illness due to ineffectiveness of case isolation as a control intervention (WHO 2014).

In all countries, the disease was widely reported across the districts but some had higher incidence than others. Therefore, the disease displayed high levels of

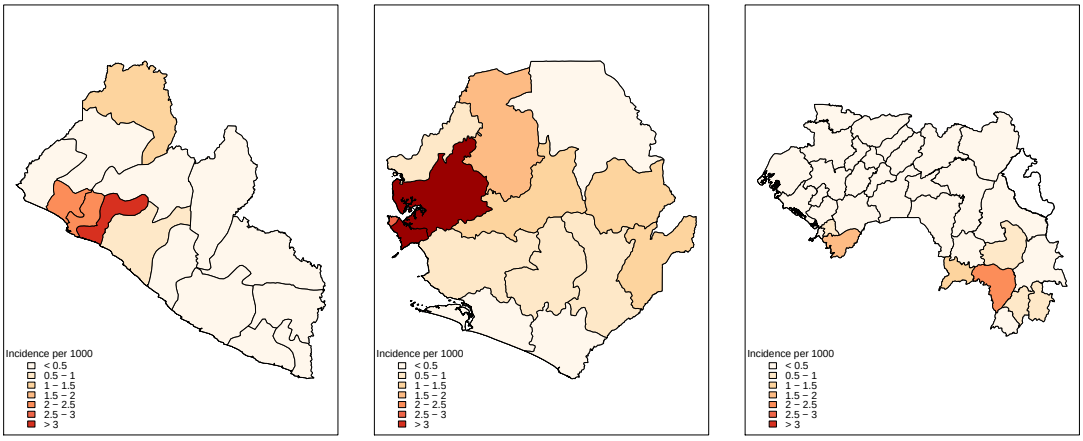

Figure 3.1: Average EVD incidence per 1000 population for the entire epidemic period at different spatial scales: (from left to right: Location of Liberia, Sierra Leone and Guinea relative to each other, incidence at county level in Liberia, districts level incidence in Sierra Leone and prefecture level incidence in Guinea

heterogeneity. Figure 3.1 shows the differences in the average incidences per 1000 population of EVD in the three countries over the whole epidemic period.

In each of the three maps, higher incidences were observed in the more densely populated urban areas than in rural areas. For example, Montserraddo county in Liberia, where the capital city Monrovia is located, recorded the highest incidence. Temporal heterogeneities in incidence at the country level are shown in Figure 3.2. However, these national epidemic curves mask the heterogeneities within the countries. Within-country heterogeneities are revealed in the epidemic curves for the sub-national units of each country; see Appendix C.

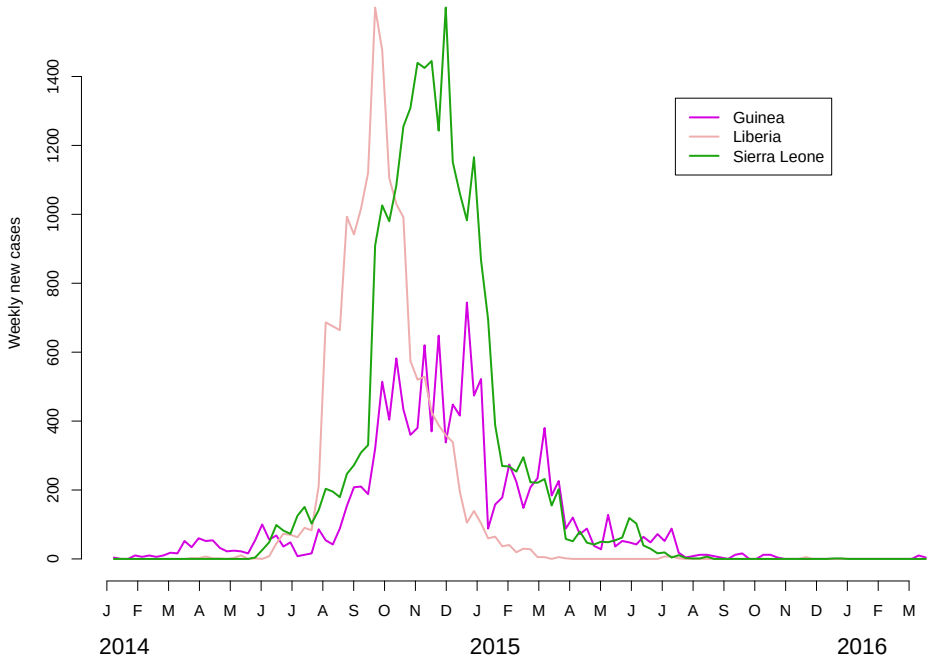

Figure 3.2: Epidemic curves for EVD for each of the affected counties showing temporal heterogeneity in disease incidence at the country level

We analysed case counts reported by the WHO as of May 2016 after the epidemic ended in all the 3 countries (*WHO* 2017). Each reported case was classified as confirmed, suspected or probable. Confirmed cases are lab-confirmed while suspected cases are based on possible links with previously known cases. For our analysis, we used the total number of reported cases (all categories). Official district population estimates for the year 2014 were used in the analysis. These data were collected from the national statistical agencies of the three countries.

The data were reported as total number of reported cases in one-week windows. To account for the incubation period and any delay in seeking care after symptom onset, we shifted back this reporting window by 17 days, corresponding to an incubation period of 11 days and an average reporting delay of 6 days. We then randomly sampled the date of infection in this one week window. This means cases reported in the given one-week window to give imputed dates of infection. To check the consistency of these imputed dates, we repeated the imputation process 100 times. The resulting epidemic curves are shown in Figure 3.3

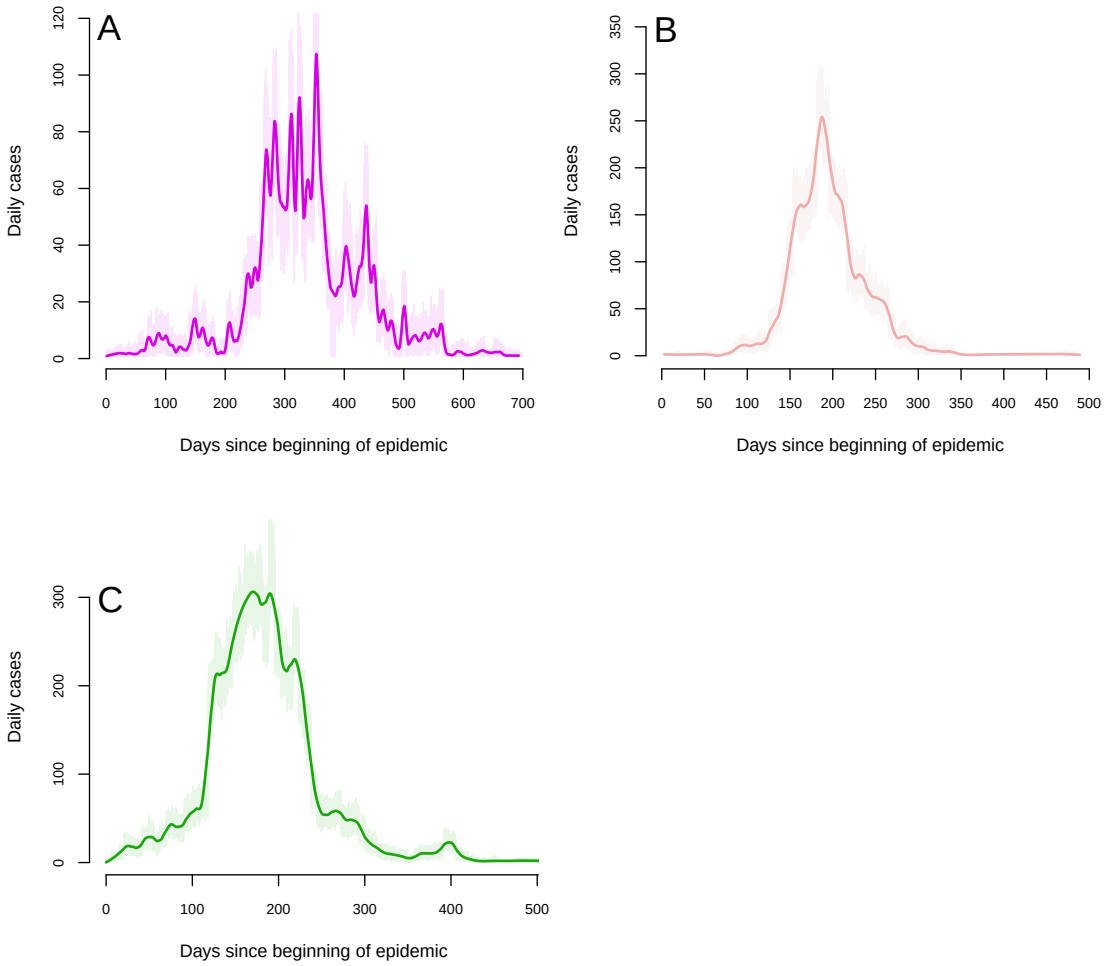

Figure 3.3: Epidemic curves for EVD epidemic in West Africa with imputed disease onset times; (A) Guinea, (B) Liberia and (D) Sierra Leone. The faint lines in each plot are calculated from the different imputations, whilst the solid lines are loess smooths of pointwise medians; Cleveland (1979)

6.2 Estimates of country-level R_t

Figure 3.4 shows country-level Wallonga-Teunis estimates of R_t in each of the three countries.

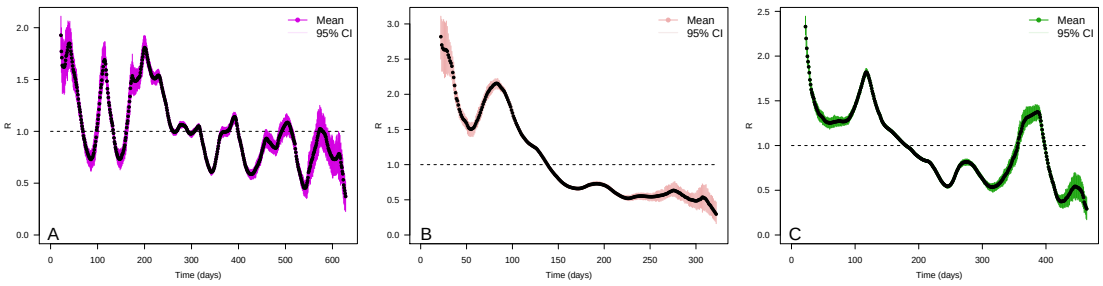

Figure 3.4: Wallinga Tuenis estimates of the non-spatial reproductive number for the 2013-16 EVD epidemic in West Africa, (A) Guinea (B), Liberia and (C) Sierra Leone

For Guinea, it can be seen that the epidemic reached its peak 10 months after the first case was reported. The reproductive number is at its highest before this peak, and thereafter takes excursions above and below 1 until the epidemic finally dies out around day 600. There were episodes when the epidemic was under control ($R_t < 1$) followed by out-of-control episodes.

In Liberia, the Ebola epidemic reached its peak almost midway through the epidemic and remained out of control for a few more weeks, returning to an in-control state ($R_t < 1$) at about 120 days. The epidemic then stayed under control, although with varying R_t, until it died down completely after about 300 days.

For Sierra Leone, the epidemic grew throughout the first few months, reaching the peak at around 120 days after the onset of the epidemic. Subsequently, R_t dropped below one until around day 350, when it rose above the $R_t = 1$ threshold for several weeks before again falling away until the epidemic ended around day 450.

In all countries, the 95% confidence intervals appear to be much narrower during the peak months of the epidemic when there is a lot of data and thus less uncertainty. On the other hand, there are observed wider confidence intervals during the early and latter parts of the epidemics. This is due to the relatively small number of cases for estimation of the R_t.

6.3 Spatial extension

6.3.1 Model parameters

Table 3.1 shows model parameter estimates for each of the fitted transmission kernels.

Table 3.1: Partial maximum likelihood parameter estimates of the power law (PL) 3.13 and lagged power law kernels for the three countries and their associated AIC values

Country	Model	Parameter	Estimate	95% CI	AIC
Liberia	PL	γ	0.87	(0.069,1.67)	-6.08
	LPL	γ	0.001	(-0.0053,0.0073)	-4.23
		δ	0.37	(-1.55,4.24)	
Sierra Leone	PL	γ	2.11	(0.17,11.61)	12.34
	LPL	γ	0.049	(-0.19,0.29)	30.18
		δ	-3.31	(-28.45,21.83)	
Guinea	PL	γ	0.47	(0.066,2.36)	-4.26
	LPL	γ	0.078	(-21.26,15.19)	179.45
		δ	-3.03	(-0.084,0.24)	

After maximizing the partial log likelihood, the partial likelihood estimates and 95% confidence intervals γ in the power law model were 0.87 (CI:0.069,1.67) for Liberia, 2.11 (CI: 0.17,11.61) for Sierra Leone and 0.47 (0.066,2.36) for Guinea.

To choose between the power and lagged power laws, we compared the values of the Akaike information criterion (AIC) for the two models in each country. In each case, the parameter σ of the lagged power law was fixed at $\sigma = 1$ while δ and γ were allowed to vary. Overall, the power law was found to fit better in all three countries as indicated by its smaller AIC values. We therefore used the power law to calculate the transmission probabilities and the spatial reproduction numbers $R(y, t)$. Figure 3.5 shows the plots for the power law kernels and their 95% confidence intervals. The confidence intervals are very wide.

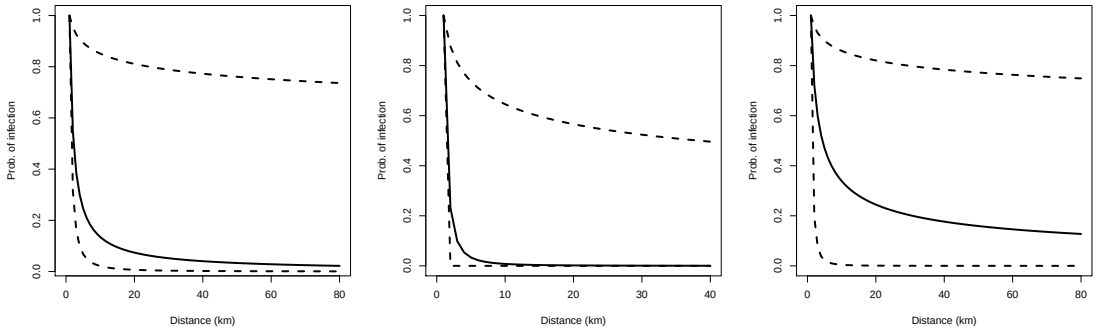

Figure 3.5: Power law transmission kernel as a function of distance for the three countries with confidence intervals, from left to right: Liberia, Sierra Leone and Guinea

6.3.2 Spatial estimates

The reproductive number estimate for a whole country without taking into account the spatial heterogeneity is given by $R(y, t)$ when y is the single spatial unit corresponding to the whole country. Compared with the WT estimates of $R(y, t)$ in Figure 3.4, some similarities in the general shape of the curve can be seen. In this section, we present only the spatial $R(y, t)$ estimates for Liberia. Using our approach, the estimates are given in Figure 3.6.

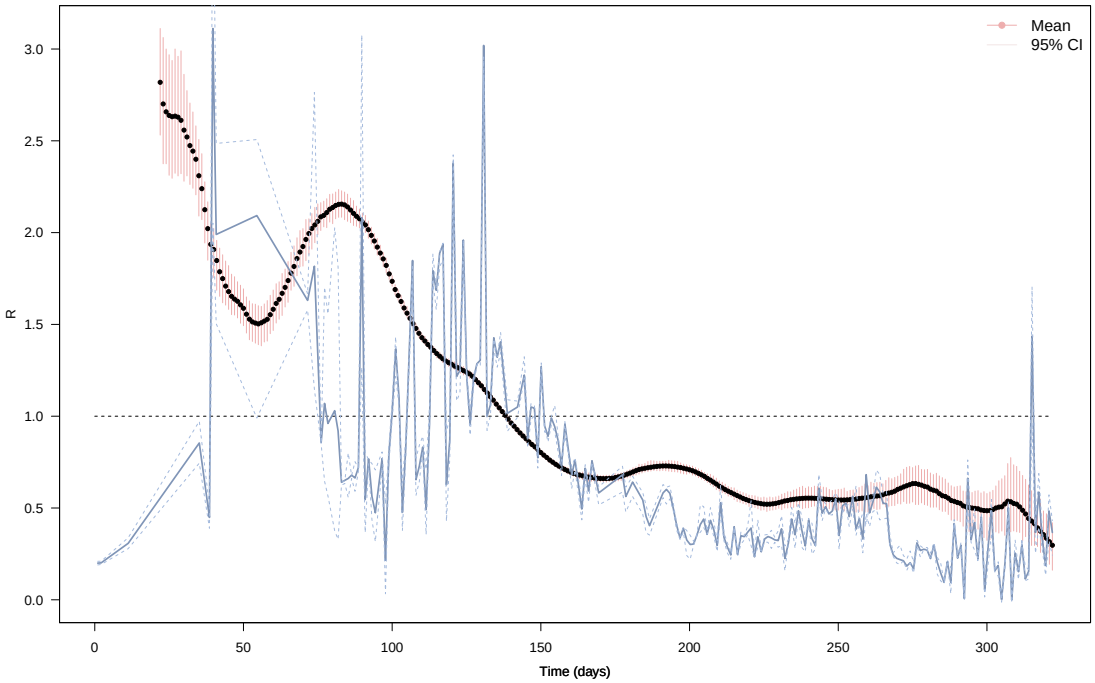

Figure 3.6: Global spatio-temporal reproductive number number $R(y,t)$ for Liberia (blue colour) superimposed over the global R_t estimate using WT method

Compared to the WT estimate for Liberia, there is a general similarity in the shape. However, the estimate of $R(y,t)$ shows the epidemic was still out of control as shown by values of $R(y,t) > 1$ for a longer period compared to the WT estimate. This could because some counties started to report more cases later on in the epidemic and their contribution to the overall R_t is taken into consideration. The confidence intervals for the two figures seem similar in the sense that they are both wide early in the epidemic before narrowing up and later widening again towards the end.

Spatial estimates $R(y,y,t)$ are reproductive numbers at time for a given location, y, that are transmitted by an infectious case in the same location at time t. Figure 3.7 shows these estimates for each of 12 Liberian counties. Some of the confidence intervals are narrow, indicating that imprecise estimates of the transmission kernel can nevertheless lead to usefully precise estimates of reproductive numbers.

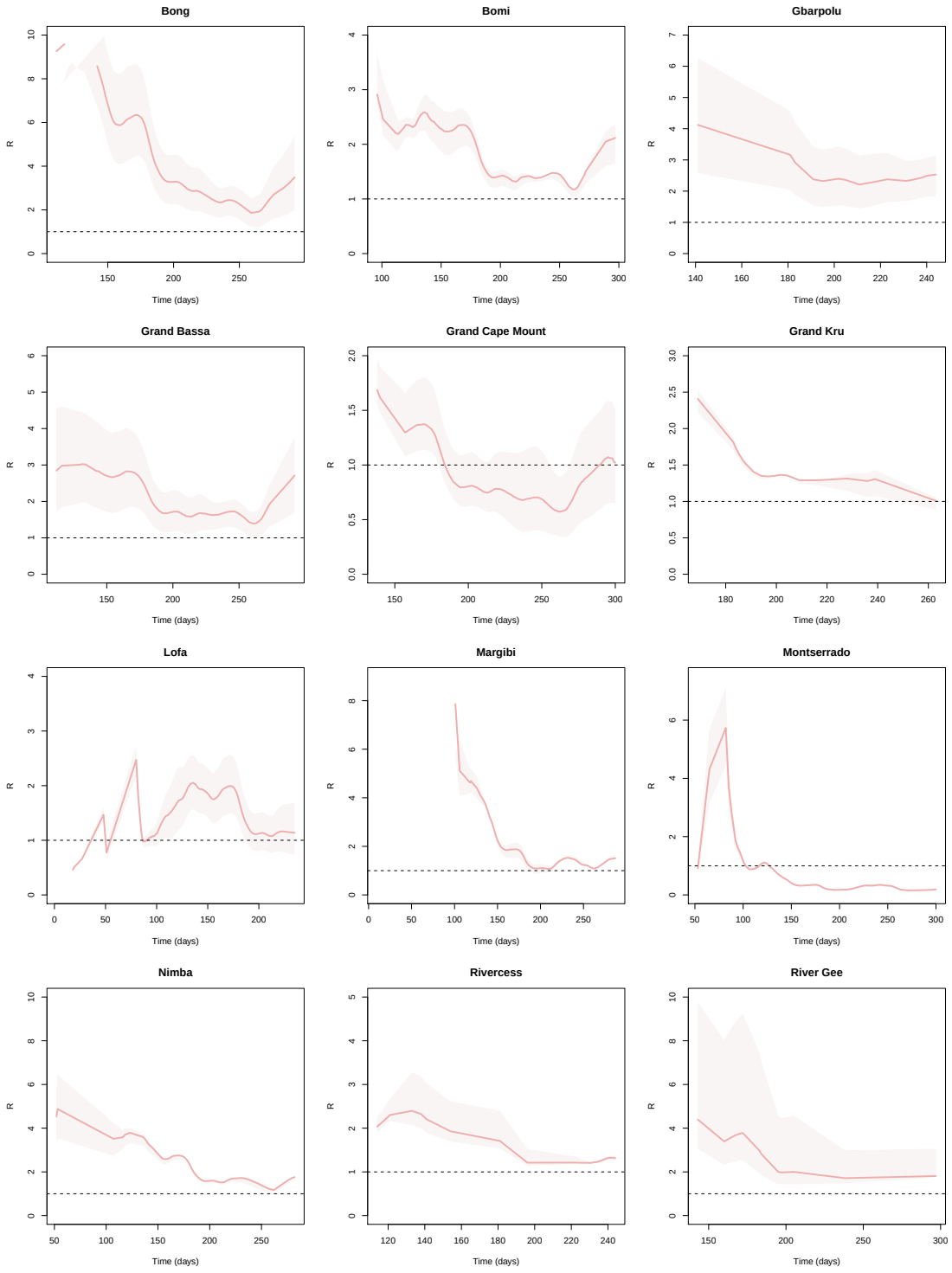

Figure 3.7: Spatial reproductive number estimates, $R(y, y, t)$ for locations y corresponding to 12 out of the 15 counties of Liberia.

In comparison, the WT estimates within the same county show some noticeable differences in shape and are shown in Figure 3.8. Of note is the fact that with the WT estimates, the overall R_t for Liberia (Figure 3.6B) closely resembles the R_t pattern for Monsterrado in Figure 3.7. Being the county with the highest burden of EVD, this county contributes the most to the global R_t estimate. When spatial

heterogeneities are taken into account, this effect is reduced as all counties now contribute to the R_t.

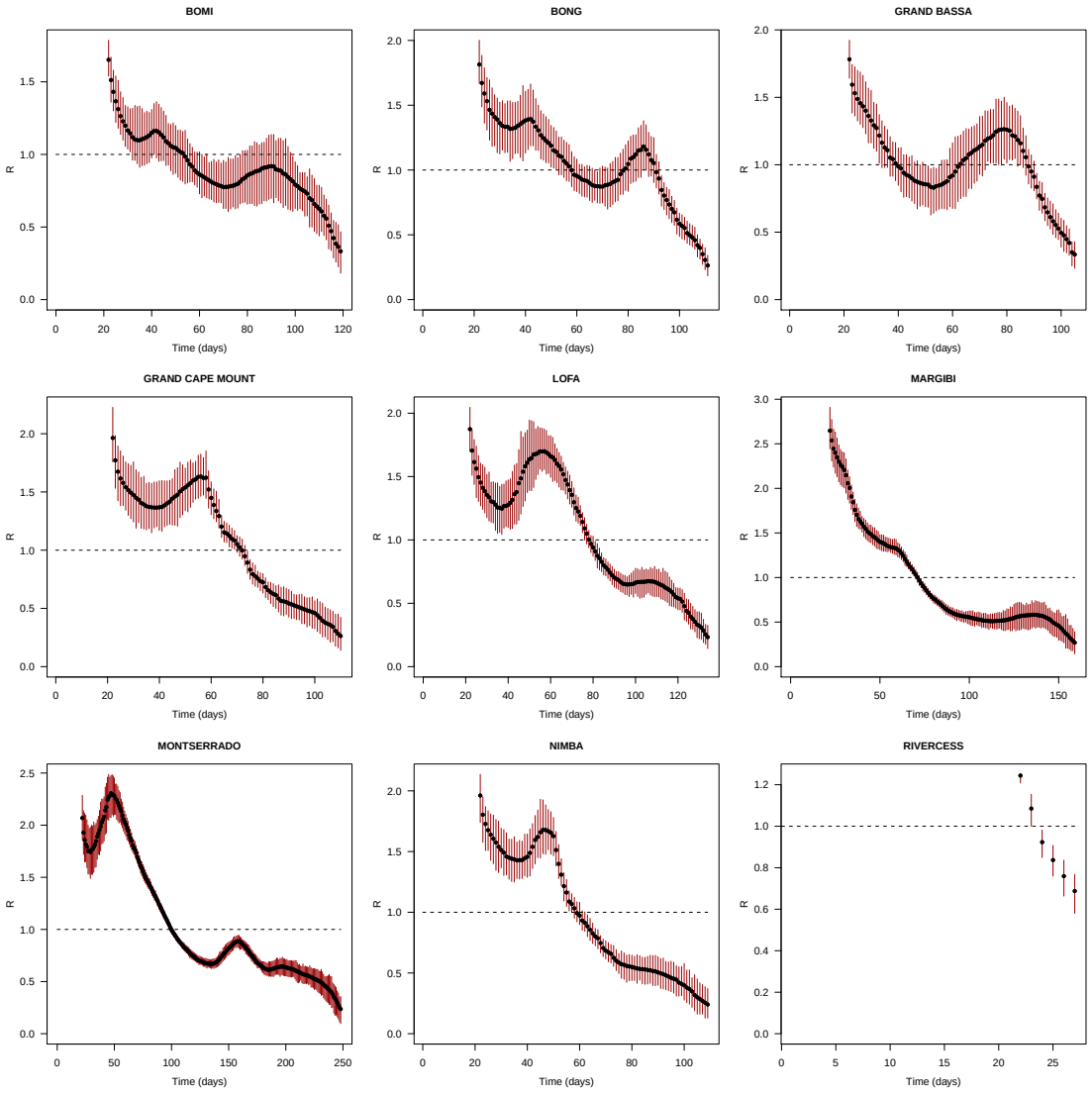

Figure 3.8: WT estimates R_t within selected counties of Liberia

Finally, we present estimates of the reproductive numbers $R(x, y, t)$ at location x transmitted by the infectious case in a different location y at time t. This paints a very different picture from the within-country reproductive numbers $R(y, y, t)$. In the counties shown in Figure 3.9, $R(x, y, t) < 1$ except for locations $y =$ Montserrado where $R(x, y, t)$ is sometimes greater than one. This could be because there is little interaction between these locations. For example, Sinoe and Maryland counties do not share a boundary, hence it is plausible that the reproductive number between them is small. The generally higher estimates of reproductive numbers from origin cases in Montserrado county indicate high levels of interaction between Montserrado and other counties. Montserrado has the highest population and contains the capital city Monrovia. Therefore, rate of

human movement from this county is likely to be both higher overall, and spatially more dispersed, than movement rates from other counties.

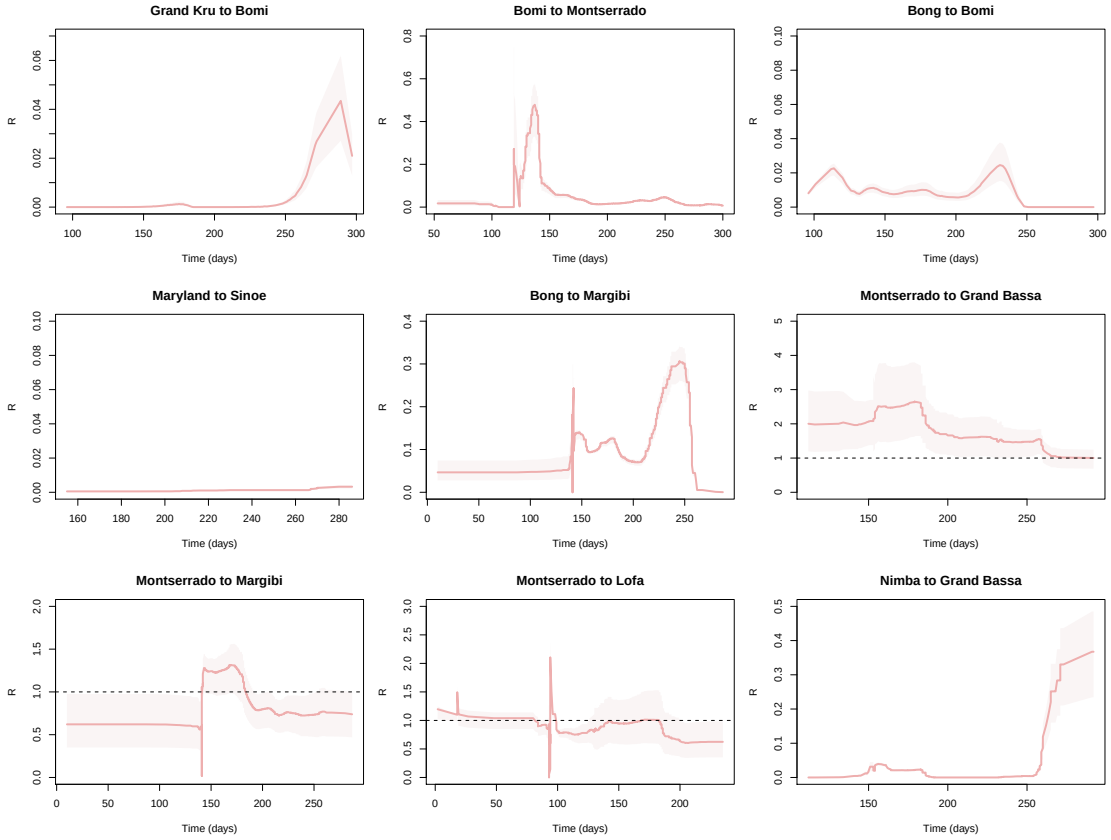

Figure 3.9: Spatial reproductive number estimates in Liberia, $R(x, y, t)$ at location x transmitted from location y at time t.

7 Discussion

The extension of the reproductive number to a spatio-temporal framework provides a more complete picture of the disease transmission process. Because R_t only looks at changes with time, it cannot provide information on any spatial factors that affect the transmission dynamics of a disease. Accounting for the spatial dimension can lead to the identification of transmission hotspots, as indicated by high values of the reproduction number estimates in Figure 3.7. The average spatial reproduction numbers between pairs of districts can also help to identify those districts that are the primary drivers of the outbreak. Constantly high values from one district to another indicate particularly important transmission routes between districts. Rarely are infectious diseases homogenous in their spread over a geographical region. Hence, our methodology will help in the design and application of interventions designed to control outbreaks.

The epidemics in the three countries show broadly similar spatial and tempo-

ral characteristics. For example, urban areas with high connectivity show high transmission potential. This indicates similar driving forces, such as demographic factors or cultural practices; the three countries share boundaries and have similar socio-demographic characteristics.

Our methodology is applicable to different infectious diseases. Highly infectious diseases with relative short incubation periods are best suited to the method, since their serial interval distribution can often be known to a good approximation. Also, reporting delays that are short relative to th corresponding incubation period are less likely to hide intermediate events in a chain of infections. In addition to Ebola, examples that meet these conditions include measles and influenza. For infections with long incubation periods, it becomes a challenge to identify a chain of infection of who infected whom, which plays a key role in the WT method and the spatial extension described here.

Parametric modelling of transmission rates between spatial locations allows for a principled statistical inference framework to be used in model-fitting. The partial likelihood method has previously been used to model disease transmission, and in other areas of application such as ecology (Diggle 2006; Diggle et al. 2010; Lawson and Leimich 2000). The partial likelihood approach is also more straightforward computationally than full maximum likelihood, at the cost of some loss of efficiency.

As is always the case, the precision of estimated quantities of interest, here the spatial reproduction numbers, improves with the richness of the available data. Typically, in an emergent outbreak only reported onset times are readily available. This applies when cases are notified through the official government health care system, when there is usually not enough time or resources to gather additional data. Individual-level and location-specific covariates are not usually of primary concern during the epidemic in the absence of local research institutions. Information of this kind must then be gathered from other sources. Smaller, more efficient organizations, such as within-country research institutes working closely with the governments, can provide the means to collect additional covariates, enabling more precise estimation of epidemiologically relevant quantities such as reproduction numbers.

In future work, we aim to extend the model to continuous geographical space to deal with studies of ongoing epidemics in which individual cases cases can be geo-located in real-time.

References

Cauchemez, Simon, Pierre-Yves Boëlle, Guy Thomas, and Alain-Jacques Valleron (2006). “Estimating in real time the efficacy of measures to control emerging

communicable diseases." In: *American journal of epidemiology* 164.6, pp. 591–597.

Cleveland, William S (1979). "Robust locally weighted regression and smoothing scatterplots." In: *Journal of the American statistical association* 74.368, pp. 829–836.

Cori, Anne, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez (2013). "A new framework and software to estimate time-varying reproduction numbers during epidemics." In: *American journal of epidemiology* 178.9, pp. 1505–1512.

Cowling, Benjamin J, Vicky J Fang, Steven Riley, JS Malik Peiris, and Gabriel M Leung (2009). "Estimation of the serial interval of influenza." In: *Epidemiology (Cambridge, Mass.)* 20.3, p. 344.

Cox, David R (1972). "Regression models and life-tables." In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187–220.

Diggle, Peter J (2006). "Spatio-temporal point processes, partial likelihood, foot and mouth disease." In: *Statistical methods in medical research* 15.4, pp. 325–336.

Diggle, Peter J, Irene Kaimi, and Rosa Abellana (2010). "Partial-Likelihood Analysis of Spatio-Temporal Point-Process Data." In: *Biometrics* 66.2, pp. 347–354.

Firth, David (1988). "Multiplicative errors: log-normal or gamma?" In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 266–268.

Fraser, Christophe (2007). "Estimating individual and household reproduction numbers in an emerging epidemic." In: *PLoS One* 2.8, e758.

Lawson, Andrew B and Petra Leimich (2000). "Approaches to the space-time modelling of infectious disease behaviour." In: *Mathematical Medicine and Biology* 17.1, pp. 1–13.

Lin, Feng-Chang and Jun Zhu (2012). "Additive hazards regression and partial likelihood estimation for ecological monitoring data across space." In: *Statistics and its Interface* 5.2.

Meyer, Sebastian, Leonhard Held, et al. (2014). "Power-law models for infectious disease spread." In: *The Annals of Applied Statistics* 8.3, pp. 1612–1639.

Shultz, James M, Zelde Espinel, Maria Espinola, and Andreas Rechkemmer (2016). "Distinguishing epidemiological features of the 2013–2016 West Africa Ebola virus disease outbreak." In: *Disaster Health* 3.3, pp. 78–88.

Wallinga, Jacco and Peter Teunis (2004). "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures." In: *American Journal of Epidemiology* 160.6, pp. 509–516.

White, Laura F, Brett Archer, and Marcello Pagano (2013). "Estimating the reproductive number in the presence of spatial heterogeneity of transmission patterns." In: *Int J Health Geogr* 12, p. 35.

WHO (2017). URL: http://apps.who.int/gho/data/node.ebola-sitrep.quick-downloads?lang=en (visited on 12/12/2017).

WHO, World Health Oganization Ebola Response Team (2014). "Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections." In: *N Engl J Med* 2014.371, pp. 1481–1495.

WHO, World Health Organization Ebola Response Team (2016). "After Ebola in West Africa—unpredictable risks, preventable epidemics." In: *New England Journal of Medicine* 375.6, pp. 587–596.

Chapter 4

A review of models of human mobility for predicting infectious disease spread

Chirombo J, Diggle P.J., Terlouw D.J., Read J.M.

4. A review of models of human mobility for predicting infectious disease spread

Chirombo J, Diggle P.J., Terlouw D.J., Read J.M.

Abstract

The spread of human infectious diseases is driven by human movement at global and local scales, through air travel, commuting, and the interaction of people. The development of a quantitative understanding of human movement patterns is, therefore, essential to effectively predict and control the rapid dissemination of diseases. The use of mobility models for understanding infection patterns is relatively recent in epidemiological studies, leading to the development and implementation of a range of ad hoc approaches when developing mobility models for infectious diseases. Yet models of human mobility describing the movement of people from one location to another have a long history of use in other disciplines. Here, we review the range of different approaches used for modelling human movement across different spatial and temporal scales, and the data types that can be used to parameterize them. We identify two main modelling approaches, namely, the gravity and radiation models. We describe the approaches taken for statistical model fitting and parameter estimation procedures for these models, and discuss their performance and highlight important limitations. Finally, we identify key challenges remaining for the formulation, parameterization and use of human mobility models within an epidemiological modelling framework.

1 Introduction

Human behaviour underpins the spread and persistence of infectious disease epidemics. Among these behaviours, movement or travel, from one location to another, is recognized to be a significant factor in the spread of diseases and their control (Wesolowski et al. 2012; Funk et al. 2010; Cliff and Haggett 2004; Grietens et al. 2015). For example, the spread of seasonal influenza has been shown to be driven by daily commuting patterns, providing connections between cities and permitting the virus to access new populations of susceptible individuals (Viboud et al. 2006). Human movement has also been shown to be a key component in the transmission dynamics of smallpox in an event that the virus is introduced into a population

(Riley and Ferguson 2006). With increasing globalization, understanding the role that human movement plays in the dissemination, dynamics and persistence of infectious diseases is crucial to the predication of future outbreak and the design of effective control solutions.

There is a rich history of studies considering human movement patterns, and the development of a quantitative understanding of those patterns, often in fields relatively far removed from epidemiology. Movement patterns have been described as a class of models known as spatial interaction models (Haynes, Fotheringham, et al. 1984). These models describe the total movement of people, as well as the flow of commodities, capital or information, from an origin location to a destination location (Sen and Smith 2012). Spatial interaction is used to describe a wide range of behaviours such as migration of individuals between locations, or the daily commuting patterns of workers.

1.1 Evidence for human mobility driving disease dynamics

Infectious diseases are responsible for high mortality and morbidity in many different parts of the world, with a particularly high health burden in developing countries (Mabey et al. 2004; Bygbjerg 2012; Dye 2014). Understanding the spread of these diseases is of great importance in the design of interventions aimed at tackling them.

Rapid improvement in travel due to increased frequency of trips and short period of time taken poses a threat to epidemic outbreaks, (Ruan et al. 2015; Chen 2015). Mass gatherings and movement of people are likely to provide a conducive environment for rapid disease spread. For example, the 2014 world cup in Brazil presented a risk for the transmission of the dengue virus among the millions of visitors that were expected to descend on Brazil for the games (Lowe et al. 2014).

Over the past few decades, there has been rapid population growth in many developing countries, and an increasing number of people living in urban environments, creating an environment suitable for the rapid spread of diseases (Neiderud 2015). With increasing urbanization, cities are providing a conducive environment where people can meet, and infectious diseases that rely on close-contact or location of hosts may be transmitted (Dalziel et al. 2013). This sharp rise in population density and urbanization has been matched by an equally marked increase in the spatial mobility of populations(Cliff and Haggett 2004). Quantifying human movement is, therefore, an important undertaking with implications in disease control and prediction, particularly in identifying the risk to otherwise unaffected regions through travel of incubating or infectious individuals. Plans to eliminate malaria in South Africa by the year 2018 exemplify this, having identified the most

effective strategy to reduce imported cases, by focusing efforts at the source of infected travelers originating in Maputo Province of Mozambique (Silal et al. 2015).

Human mobility also assists the dispersal of directly-transmitted pathogens over large geographical areas (Stoddard et al. 2009). For example, in the case of diseases such as malaria and dengue, the mosquitoes which act as vectors are known to have a limited range of migration Kaufmann and Briegel (2004). Therefore, human movement plays a key role in the transmission of the disease within and between communities (Harrington et al. 2005). In Pakistan, human mobility is responsible for the emergence of new dengue epidemics in previously low-risk areas (Wesolowski et al. 2015c). In Ethiopia, a recent study has found seasonal migrant workers to be at particularly high risk of contracting malaria (Schicker et al. 2015). Similar observations about migrant communities were also made in Myanmar (Wai et al. 2014).

1.2 Aims of the review

The overall aim of the review is to provide an overview of mobility models, their development, and current status in epidemiological modelling. Specifically, it aims to provide a summary of the main classes of models for human mobility; review parameter estimation methods used in mobility models and review the application of mobility models in infectious disease modelling.

1.3 Search criteria

We used a narrative literature review method to summarise and critique the literature about human mobility modelling. The objective of this review was to provide an overview of mobility modelling, highlight their application in different fields particularly in infectious disease epidemiology and identify gaps for further research. Therefore, a subset of articles that tackle these objectives at a detailed level was selected and reviewed to achieve a fair, balanced and comprehensive overview. Through this review, it is hoped that human mobility's role in infectious disease spread will be sufficiently understood.

1.3.1 Limitations of the search strategy

For this review, a systematic approach was not used. Therefore, some important papers may have been missed. This review is not exhaustive or definitive. While a systematic review would have been preferable, the wide application of mobility models in different disciplines would have yielded a broad area to be covered. As a

result, the current review was primarily driven by the objective to highlight their application in infectious disease modelling which may have introduced some bias.

2 Measuring human movement

There exist a variety of methods to measure and characterise human movement patterns. These range from purposive large-scale surveys, small scale tracking of individuals, to secondary analysis of information collected for other purposes, such as mobile phone call data. Here we review the major methods in the literature and discuss their limitations.

2.1 Surveys and censuses

In many countries, information on mobility is often captured through censuses or other large-scale surveys of the population. Censuses typically conducted once a decade, can capture seasonal or longer-term changes in dwelling place as well as migration patterns. The Malawian census captures movement across districts and regions over the previous 10-year period; the census records the change of residence by asking respondents about their residential history. Some censuses and large-scale surveys also capture information about home and work place locations, and so can provide information on movement between origin and destination locations (for example, workflow data derived from the United Kingdom 2011 census (*UK Data Service* 2018). Such measurements may, of course, suffer from recall bias when respondents provide that information. Survey data generally captures individual-level information, enabling mobility by different demographic groups to be assessed.

One of the challenges of mobility information derived from censuses is timeliness. Since most censuses are conducted every ten years, studying inter-censual movement patterns is impractical. Also, census information typically fails to capture very fine temporally resolved data, often being limited to cross-sectional information collected within a short time period. The inherent high costs of carrying out comprehensive and representative surveys pose another challenge. Consequently, mobility questions are usually incorporated into other household surveys such as demographic and health surveys (DHS). Where this occurs, questions related to mobility tend not to be very detailed in their coverage (Wesolowski et al. 2014b).

2.2 Mobile phone records

There has been a development in measuring human mobility at the population scale using mobile phone data (Gonzalez et al. 2008; Chen et al. 2018; Jiang

et al. 2017; Williams et al. 2015). Whenever an individual makes a mobile phone communication (voice call, short message service (SMS), data, allowance top up, electronic payment, etc), the senders and receiver's identities, date and location of the communication are recorded by the mobile companies. Their location can then be approximated by the nearest cell tower through which the communication was routed. By gaining the anonymized call detail records (CDRs) together with the coordinates of all the cell towers, it is possible to estimate the location of phone users. Individual travel trajectories may then be mapped over a specified period. The ever-increasing use of mobile phones, even in developing countries, has made it possible to measure human movement in many settings (Buckee et al. 2013).

Unlike census data, mobile phone data provides information on regular movement patterns at different spatial and temporal scales, (Wesolowski et al. 2012). These data make it possible to follow population changes over a short period such as a week or a season (Deville et al. 2014). This approach offers additional benefits compared to surveys, as the data collection may be automated, and routinely recorded by service providers.

CDRs are increasingly being used to estimate human mobility patterns across the globe for different purposes. However, the privacy of subscribers is a concern when using mobile phone data (De Montjoye et al. 2013). This problem may be addressed by ensuring data released by the mobile companies for research purposes are anonymized, typically through the aggregation of information, yielding information on the flows of individuals between locations rather than the spatial trajectory of individuals. Additional challenges in using the data include storage, representation, analysis, and computation complexity (Asgari et al. 2013). Phone sharing, common in developing countries, and a bias towards wealthy urban males are further challenges in using mobility information derived from mobile phone data (Blumenstock and Eagle 2010). Finally, the absence of individual level information (e.g., demographic information or health status) excluded due to privacy concerns, or often not recorded by the service provider makes it extremely difficult to classify mobility patterns by some key demographic characteristics known to be important for disease transmission (such as age).

Despite these challenges, mobile phone data have been used for a wide range of applications, particularly in developing countries where comprehensive datasets on mobility patterns from other sources are not usually available. Bengtsson et al. (2015) used CDR data for predicting the spatial spread of cholera. In Kenya, they were used to map net malaria exporting and importing locations for more targeted disease control (Wesolowski et al. 2012) and in Namibia for mapping malaria risk for elimination interventions planning (Tatem et al. 2014). Mobile phone data have also been used in Kenya to quantify seasonal movement patterns driving

rubella disease transmission dynamics (Wesolowski et al. 2015a). During the West Africa Ebola epidemic, these data were also utilized to understand the spread of the disease (Wesolowski et al. 2014a). They have also been used to predict the spread of dengue epidemics in Pakistan (Wesolowski et al. 2015c).

2.3 Other methods to measure mobility patterns

A number of additional methods have been used to capture movement patterns. For example, Global Positioning System (GPS) enabled devices have been used to track the movement of individuals (Searle et al. 2017). GPS tracking is a common method that can record with high accuracy the location of the device (and the participant) every few seconds. This has been utilized in vehicles, mobile phones as well as dedicated devices (such as navigational aids). While GPS-enabled smartphones are increasingly widespread, more recently the development of wearable fitness trackers can also provide another source of location tracking (Meekan et al. 2017).

GPS tracking, however, has some limitations including data loss due to signal drop-out, dead batteries and misuse of the device (Krenn et al. 2011). Some other technical challenges such as signal noise, signal obstruction inside buildings are also common limitations to GPS tracking (Paz-Soldan et al. 2014).

Social media is another source of mobility data that is increasingly being used. With billions of people using social media platforms such Facebook (*Facebook* 2018), Twitter (*Twitter* 2018) and Weibo (*Weibo* 2018), it is possible to identify individual trajectories based on georeferenced posts users can make. Other opportunities for capturing location and movement data include geotagged pictures taken by mobile phone users. Photo sharing services such as Flickr (*flickr* 2018) provide georeferenced pictures that have been used to infer mobility patterns (Zheng et al. 2012).

3 Models of human mobility

At best, information collected by the previously outlined methodologies can describe a representative sample of entire population. Epidemic modelling is often concerned with predicting infection risk and transmission events for the entire population under study. Consequently, developing models of human mobility offers a way to predict movements within the whole population, and to describe how movements may impact disease dynamics and control. Such mobility models, therefore, provide a way of quantifying the amount of travel between locations when direct observation is not feasible. Two main model formulations have been proposed: the gravity model (and derivates) and the radiation model. Both models attempt to describe the flow

of people (though can also be applied to other entities or commodities) between an origin location and destination location and relate this flow to properties of those locations. Population density is the most common location property considered, though some studies do consider alternatives. We discuss each model formulation in turn.

3.1 The gravity model

The earliest known formulation of a model of human population mobility was made in the 19th century by HC Carey (Carrothers 1956). In 1885, the concept was later used to explain migration movements (Ravenstein 1885). Ravenstein noted that migration movement tends towards cities with large populations and that volume decreases with the distance between origin and destination. The concept was later generalized in the early 1940s by Zipf and Stewart (Carrothers 1956). These interpretations had a common formulation within their models, subsequently termed the 'gravity' model of mobility.

The gravity model, borrowed from Newton's law of gravitation force, describes the total flow of subjects between any two distinct locations while considering the distance between those locations. Here, subjects may be human individuals, or commodity units. Mathematically, it is typically represented in the form

$$T_{ij} = \frac{P_i^{\alpha} P_j^{\beta}}{d_{ij}^{\gamma}}, \tag{4.1}$$

where T_{ij} is the total flow of subjects from origin i to destination j during a specific period. P_i and P_j are the total population sizes at the origin and destination locations i and j respectively, and d_{ij} is a measure of the distance separating the two locations. The parameters α, β, and γ are unknown and usually obtained from fitting a linear regression model to the data (Wesolowski et al. 2015a).

Not all gravity models use population size, however, as the sole property of locations. Population density is often employed for models where space is considered as a discrete grid. Gravity models of international trade use information on gross domestic product (GDP) of the two trading countries instead of population size (Chaney 2018; Fratianni 2007; Bergstrand 1985)

These parameters are employed differently in various interpretations of the gravity model, with some formulations disregarding them altogether. They have been applied to the origin and destination populations to account for hidden (unobserved) variables that may be specific to the local regions (Toole et al. 2015). For example, α can quantify a 'push' or repulsion effect at the origin, while β describes a 'pulling' or attraction effect of the destination location. In this interpretation, destinations with large populations have a stronger pull effect

given by a higher value of β. This assumption is often justified by considering the population size to be a proxy for work opportunities for commuters, or sales opportunities for commodities. The parameter γ is the distance friction coefficient (Chen and Xu 2013) and indicates the ease of connectivity between locations. For example, two areas separated by a range of mountains may have a higher value of γ compared to those connected by a fully surfaced tarmac road and without geographical obstacles between them. There is an assumption in spatial interaction models that distance functions capture the spatial dependence between origin and destination pairs (LeSage and Pace 2008; Tsutsumi and Tamesue 2012). However, this assumption has been challenged by others (Porojan 2001). The residuals from the models showed spatial dependence hence the need to properly account for the spatial dependence in the model. By tuning the values of the gravity model parameters accordingly, the amount of flow will also be affected. Furthermore, the parameter γ is assumed to be uniform for all locations pairs despite possible differences in the strength of connectivity between pairs of locations.

At an international scale, the gravity model can be augmented by adding covariate information such as whether countries share boundaries (Anderson and Van Wincoop 2003). The distance term in the model can be defined in terms of factors that provide resistance to trade (Silva and Tenreyro 2006). In gravity models, there is friction that is provided by the competing destinations for the flow emanating from the origin i.

3.1.1 Conservation constraints of the gravity model

The gravity model presented by equation 4.1 has its own deficiencies. Wilson (1967) pointed out the model is unbalanced in that when both P_i and P_j are doubled, for example, the total flow quadruples instead of doubling also. In the gravity model, there are some constraints that may be enforced. The total flow between the origin i and destination j should equal the total flow originating at the origin i, $\sum_j T_{ij} = P_i$. Likewise, it must be equal to the flow terminating at the destination location j, i.e. $\sum_i T_{ij} = P_j$. To achieve that, Wilson (1967) introduced constants A_i and B_j for the origin and destination respectively. A gravity model with constraints and a general function for the distance was defined by Wilson (1967) as follows

$$T_{ij} = A_i B_j O_i D_j f(d_{ij}) \tag{4.2}$$

where T_{ij} is the total flow, O_i is the total number of individuals leaving the origin and D_j is the total number arriving at the destination. A_i and B_j act as balancing factors designed to ensure that the conservation constraint is achieved (Dennett 2012; Grange et al. 2010). The function $f(d_{ij})$ is a measure of impedance and

can depend on the distance between i and j, on travel time, on the cost of travel, or a weighted combination of some factors (Wilson 1967). This function can be specified in different forms such as the exponential decay function.

When the origin is adjusted by the correction factor $A_i = \frac{1}{\sum_j D_j f(d_{ij})}$, an origin constrained gravity model is obtained (Wilson 1967). This makes sure that the number of trips produced at the origin cannot exceed the number of people. In a destination constrained model, the balancing factor $B_j = \frac{1}{\sum_i O_i f(d_{ij})}$ is applied to the destination so that the total flow does not exceed the number of opportunities on offer (MacLachlan 2011). When either of these constraints have been applied, a singly-constrained gravity model is produced. When both have been applied in a model, a doubly-constrained gravity model is the result.

3.1.2 Variations of the gravity model

An area of active research over recent years in spatial interaction modelling has been the formulation and estimation of gravity models, resulting in different forms being used. One of the earliest modifications to the gravity model is the linearization process whereby the model is transformed into a linear model following the generalized linear model (GLM) framework (Ewing 1974). Parameter estimation in the traditional gravity model formulation can be computationally intensive hence the development of more efficient ways of estimating the parameters. Computational limitations also drove research into alternative forms of estimating gravity models. Presenting the gravity model as a linear model is appealing as it makes parameter estimation and interpretation straightforward. As a result, the log-linearized form has been used several times, despite well-documented limitations, including bias introduced by the log transformation, heteroscedasticity, and sensitivity to zero flows between origin and destination (Burger et al. 2009). Heteroscedasticity is caused when the assumption of constant error terms for all origin and destination pairs is violated. Furthermore, log-linearization can lead to the misspecification of the function form leading to unreliable and imprecise gravity type models (Fik and Mulligan 1998).

Equation 4.1 can be log-linearized as follows

$$\log(T_{ij}) = \alpha \log(P_i) + \beta \log(P_j) + \gamma \log(d_{ij}) \tag{4.3}$$

Covariates that are known to affect the flow can then be added to the model in the usual manner. For example, Wesolowski et al. (2015a) have incorporated the proportional of the population that was male, and constraints on the number of trips between locations as additional covariates. A gravity model implemented as a regression model with covariates has been applied in West Africa to investigate the

role of socio-demographic and environmental factors on inter-provincial migration (Henry et al. 2003). A further study has developed a Bayesian hierarchical model for human mobility that accounts for attraction and repulsion effects across the locations, as well as allowing for correlation between how attractive an area is and how retentive it is, restricting flow from its locations (Congdon 2010). This model is an improvement over the fixed effects gravity regression models that assume independence between the origin and destination effects. The Bayesian framework of the model makes the estimation of random effects easier.

As the model outcome variable is count data of individuals moving between locations, where the flow between i and j is a non-negative value, the Poisson regression model is a natural GLM to consider for modelling the mean flow,

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \log(P_i) + \beta_2 \log(P_j) + \gamma \log(d_{ij}) \tag{4.4}$$

where β_1 and β_2 are regression coefficients. This GLM has the desirable characteristic that flow can be interpreted in terms of probabilities taking all the properties of Poisson regression such as risk ratio interpretation of coefficients. The decay parameter γ would normally take a negative value.

An extension to the Poisson regression model includes additional terms of attractivity/retentivity and an accessibility index, A_j:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \log(P_i) + \beta_2 \log(P_j) + \gamma \log(d_{ij}) + \delta \log(A_j) + s_{1i} + s_{2j} \tag{4.5}$$

where $A_j = \sum_{r \neq j} \frac{p_r}{d_{rj}}$ is the competing destination index that measures the proximity of destination location j to alternative destinations and thus captures people's propensity to choose one location over the other (Congdon 2010). P_r is the population at alternative destination r while d_{rj} is the distance separating destination j and any alternative destination r. Large values of A_j indicate the destination is closer to other alternative destinations with large populations, while a small value of A_j is indicative of a spatially isolated destination (Pellegrini and Fotheringham 2002). s_{1i} and s_{2j} are push and pull scores at the origin and destination respectively which are treated as spatially correlated random effects in the model (Congdon 2010). Care should be exercised, however, when deciding on the functional form of the model to be adopted. Fik and Mulligan (1998) recommend that the appropriateness of an adopted functional form should be scrutinized to avoid misspecification. For example, the common practice of rearranging gravity type models into a linear form to simplify estimation and interpretation may be inappropriate (Fik and Mulligan 1998). They provide evidence that the estimation and evaluation of gravity type models are sensitive to the data transformation and use of a flexible functional form.

3.1.3 Transformations of gravity models

As a safeguard against misspecification, Kau and Sirmans (1979) suggested that the functional form should be determined from the data and not specified *a priori*. A method for finding the optimal form is the use of Box-Cox transformation (Box and Cox 1964). The Box-Cox transformation is defined as;

$$y(\theta) = \begin{cases} \dfrac{(y^\theta - 1)}{\theta}, & \text{if } \theta \neq 1 \\ \log y, & \text{if } \theta = 0 \end{cases}$$

and is aimed at ensuring that the usual assumptions for the linear model, $y \sim N(X\beta, \sigma^2 I_n)$ is met. When the gravity model has been log-linearized, the Box-Cox transformation can be employed to bring the transformed gravity model into the proper log-linear model specification for counts by varying the Box-Cox transformation parameters. In the gravity model setting, the transformations are aimed at generating a general flexible functional relationship which captures the underlying relationships between model components.

In real life applications, a functional form may best be chosen empirically based on the relationship between the dependent and independent variables. Physical laws can also dictate the form to be adopted (Box and Cox 1964). For instance, depending on some relationship in the problem under consideration, such as $y \propto x_1^{\phi_1}, \ldots, x_p^{\phi_p}$, a linear relationship may not be appropriate hence the need for transformations. Consider the multiplicative gravity model,

$$T_{ij} = K P_i^\alpha P_j^\beta d_{ij}^\gamma \mu_{ij} \tag{4.6}$$

with the proportionality constant K. As in the usual linear model framework, the random error term $\log(\mu_{ij})$ is assumed to be normally distributed with constant variance as well as being independent (Fik and Mulligan 1998). When expressed in a log-linearized method, the model is

$$\log(T_{ij}) = \log(K) + \alpha \log(P_i) + \beta \log(P_j) + \gamma \log(d_{ij}) + \log(\mu_{ij}) \tag{4.7}$$

In this formulation,$\gamma < 0$ to capture the expected decay with distance while $\alpha > 0$ and $\beta > 0$ capture origin and destination effects. After the Box-Cox transformations are applied, the model is specified as

$$\log\left(T_{ij}^\theta\right) = \log(K) + \alpha \log P_i^{(\phi)} + \beta \log P_j^{(\phi)} + \gamma \log d_{ij}^{(\phi)} + \log(\mu_{ij}) \tag{4.8}$$

For the response variable, the transformation is

$$T_{ij}^{\theta} = \begin{cases} \dfrac{(T_{ij}^{\theta} - 1)}{\theta}, & \text{if } \theta \neq 1 \\ \log T_{ij}, & \text{if } \theta = 0 \end{cases}$$

while for the x_jth independent variable (i.e. P_i, P_j and d_{ij}) is

$$x_j^{\phi} = \begin{cases} \dfrac{(x_j^{\phi} - 1)}{\phi}, & \text{if } \phi \neq 1 \\ \log x_j, & \text{if } \phi = 0 \end{cases}$$

θ and ϕ are the Box-Cox transformation parameters for the dependent and independent variables respectively. The values of the parameter ϕ capture the appropriate functional relationship to be determined together with the other regression parameters. After the transformation, no restrictions are required to be applied to the terms in the model. As a result, the underlying relationship dictated by the physical laws, for example, is preserved. For this reason, the Box-Cox transformations can capture the optimal functional forms which would otherwise have been misspecified. With the correct Box-Cox parameters, the transformed model can convert readily to the restricted untransformed model. For example, when $\phi = 0$, the model changes to the log-linear model.

3.1.4 Parameter estimation for gravity models

Several estimation methods for gravity models ranging from linear to nonlinear methods have been proposed. LeSage and Pace (2008) extended the linear models to take account of the spatial connectivity of the regions. Ordinary least squares (OLS) is a widely used estimation method in scenarios where the gravity model has been log-linearized. Both regression coefficients and the Box-Cox parameters can be estimated by maximizing the likelihood function. Furthermore, one can statistically test the appropriateness of transformations by making use of the likelihood ratio test where the restricted model is the null model and is compared to the transformed model. A significant likelihood ratio test then leads to the adoption of the transformed model. These different estimation methods can broadly be classified under statistical and economic approaches depending on how widely used they are in a field. Since gravity models have been studied in economics for a longer time than in epidemiology, estimation methods are more developed. Methods such as Poisson Quasi-maximum likelihood (PQML) are widely used in the economic literature (Arvis and Shepherd 2013). Bayesian methods have also found considerable use, for example (Chakraborty et al. 2013; LeSage and Llano

2016)

The log-linearization process may lead to failure in some cases. For instance, it fails when zero flows between locations are observed, as the log of zero is undefined. Another challenge encountered when the OLS method is used for parameter estimation is heteroscedasticity, distorting the model output. In the presence of heteroskedasticity, OLS approach does not perform very well which results in the elasticities yielding misleading results. When used to model bilateral trade, OLS was found to consistently overestimate effects of measures, such as geographical distance and historical ties between countries. To surmount this issue, a pseudo maximum likelihood (PML) can be applied to the multiplicative form of the gravity model (Silva and Tenreyro 2006); this was found to perform better than the OLS. Furthermore, it provided a method to handle zero flows between origin and destination, a common feature of human mobility at larger scales. Silva and Tenreyro (2011) further validated the PML estimator and they found it to be better than the OLS even in situations where there is a high proportion of zero flows. Siliverstovs and Schumacher (2009) compared the PML applied to the multiplicative form of the gravity model with the OLS estimator applied to the log-linearized version under disaggregate trade flows which also showed the PQML performs better than the OLS under different scenarios.

Other parameter estimation approaches have been developed to solve challenges that result from the inadequacy of OLS when applied to log-linearized variants of the gravity models. Methods such as Nonlinear Least Squares (NLS) for estimating the gravity model in its multiplicative form, Feasible Generalized Least Squares (FGLS), Gamma Pseudo Maximum Likelihood (GPML) of Manning and Mullahy (2001) have been developed. Martınez-Zarzoso (2013) formally tested these different estimators and found the GPML to perform well, in some cases better than the PQML, and recommended that tests be made to select the best estimator. The results were not conclusive about the best estimator between PQML and GPML, and therefore it has been suggested that model selection tests on observed flow data are conducted on a case by case basis in order to select the best model (Martınez-Zarzoso 2013). However, it is not clear how well these approaches can perform in epidemiological applications.

3.1.5 Limitations of gravity models

Despite their increasing use, there are several challenges facing the development and application of gravity models of human mobility. For example, the data needed to parameterize the model are generally not available in most areas where there are no observations suitable for model fitting (Chowell et al. 2016). Another challenge concerns the issue of scale. Gravity models have been found to yield

different results at different spatial scales, thus challenging the assumption that they are suitable for universal application. Careful tuning for each application and its specific spatial scale has been suggested to address this issue (Yang et al. 2014). There is uncertainty as to the exact form for the impedance function in the gravity equation. Since the observed decay in interactions between people with increasing distance is typically nonlinear, capturing this decline is crucial for correct prediction from gravity models. Determining the shape and parameters of this function poses a challenge (Halás et al. 2014). Some of the candidate functions for the distance between the origin and destination include a simple Euclidean distance, the exponential function and the inverse power law (Balcan et al. 2009). This wide range of choice has the potential to affect the quality of the model and hence its usefulness in different settings. A formal model selection approach is needed to resolve this issue. Currently, it is not clear under which conditions different forms of the impedance function should be used. In the original gravity law, the inverse power law is used, but this formulation has been found to lack a theoretical basis (from a sociological perspective).

Another challenge to model spatial interactions concerns the definition of attraction measurement. It has been suggested that the attraction measure is a function of the flow between the origin and destination rather than the flow itself (Chen and Xu 2013). The gravity model also fails to account for any opportunities in locations lying between the origin and the destination. For example, the presence of a highly dense population between an origin and putative destination may be expected to affect the flow between the origin and destination (Simini et al. 2012).

Gravity models also rely on parameters that must be estimated from observed data. In the absence of such data, model fitting is impossible, and parameters fitted to other populations are usually used. The presence of collinearity among the gravity parameters also makes their estimation challenging (Xia et al. 2004).

3.2 The radiation model

The radiation model (RM) proposed by Simini et al. (2012) provides an alternative approach to the gravity model to describe human movement between locations. Unlike the gravity model, the RM is based on the intervening opportunities model of Stouffer (1940) and thus takes account of the populations between origin and destination locations. The RM attempts to address some of the limitations associated with the gravity model, and also borrows from the concepts of radiation and absorption in physics. As in the gravity model, the populations at the origin and destination act as masses. In the RM, individuals move from an origin at a rate proportional to the population and are 'absorbed' by other locations at the rate

proportional to their respective populations - absorption by a location defines an individual's destination. The probability of an individual arriving at any location, therefore, depends on their probability of not being absorbed prior to arriving there (Toole et al. 2015). The model takes the form

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}, \tag{4.9}$$

where T_{ij} is the total flow from location i to j, T_i is the total flow emanating from the origin i, m_i and n_j are the population sizes at the origin and destination respectively. Note, the model is described by Simini et al. (2012) as 'parameter free', though T_i is needs to be estimated or provided for the model to work. Lastly, s_{ij} is the total population size contained within the circle of radius r_{ij} centered at origin i, excluding the populations at locations i and j. This model has its conceptual origin in describing the job-hunting market, and is based on the premise that individuals are likely to opt for locations that maximise better job prospects and closeness to their homes. The model assumes that the number of opportunities in a location is proportional to its population size. Therefore, individuals are more likely to find a job in a location with a high population and be 'absorbed' by that location. On the other hand, if the area around the origin i has a high population, an individual would be able to find a location with good job opportunities before reaching a putative destination location j further away; the s_{ij} in the radiation model accounts for this intervening population. As such, the radiation model does not rely directly on the distance between the origin and the destination *per se*. Instead, it depends on the opportunities available between the two locations. It is, therefore, more robust to the misspecification of the impedance function than the gravity model. Its parameter free nature also avoids the model fitting challenges faced by gravity models. Despite this, it has been shown to perform better than the gravity model in certain applications (Simini et al. 2012).

One significant challenge with the application of the radiation model is the value of the outflow term, T_i which is unspecified. Hence applying the radiation model at the spatial scale of interest may be a challenge due to this missing information. In their paper, Simini et al. (2012) applied the radiation model to USA commuting data in different states, where census derived information on outflow rates are known.

The radiation model has not been widely used to date in infectious disease modelling due to the challenges in how it can be used to model epidemics (Roberts et al. 2015). In a recent study, the radiation model was used to understand mobility patterns to aid application of interventions for malaria elimination (Marshall et al. 2018). It has also been used for understanding the role of human mobility

in the transmission of Schistosomiasis in Burkina Faso (Perez-Saez et al. 2015). Other applications of the radiation model for modelling infectious disease dynamics include (Dalziel et al. 2013; Tizzoni et al. 2014).

4 Application of spatial interaction models in epidemic modelling

Mathematical models of infectious disease modelling are commonly used to inform public health policy, from improving an understanding of the mechanisms driving epidemic processes to identifying the optimal intervention for controlling the disease. Owing to their central role in determining the spatial spread of disease, it is important that the mobility models which describe movement are of a suitable model structure, parameterized and validated, to provide a relevant and accurate description of movement.

Jandarov et al. (2014) and Xia et al. (2004) applied the gravity model to study the epidemic coupling for measles through gravity time series SIR model. The gravity model for city to city contacts was also used in transmission models to derive the force of infection of pandemic influenza in cities in England, Wales and the US (Eggo et al. 2011). In Vietnam, age-structured gravity models were incorporated with SEIR models to simulate the spread of influenza between cities (Boni et al. 2009). Gravity models have also been applied to estimate the spread of vector-borne diseases (Barrios et al. 2012) and predict the global spread of influenza (Li et al. 2011). Sarzynska et al. (2013) linked gravity and metapopulation models for the spread of dengue fever in Peru. Balcan et al. (2009) applied mobility models to investigate the shaping of spatio-temporal pattern of global epidemics.

Commuting data from surveys and censuses have been used to model human movement in epidemic models. Charaudeau et al. (2014) used district level commuting data derived from census to show that commuting is correlated with influenza like illnesses in France. In South Korea, Lee et al. (2018) used a metapopulation model linked with commuting flow to study the spatio-temporal pattern of the spread of H1N1 influenza in 2009. District level commuting and case data were aggregated to the regional level to investigate patterns of disease spread. Dalziel et al. (2013) used individual commuting data from the 2006 Canadian census to investigate the differences in human mobility and its relationship with epidemic dynamics. The observed differences in mobility between cities was found to be enough to cause differences in epidemic dynamics among cities (Dalziel et al. 2013). Riley and Ferguson (2006) used UK census-derived commuting data to describe the spatio-temporal dynamics of smallpox if it were introduced in the community.

In addition to the commuting data, mobile phone data has also been used in epidemic modelling. Finger et al. (2016) developed a model incorporating human mobility information derived from mobile phone CDRs to investigate role of mass gatherings on the spread of cholera in Senegal. In Haiti, CDRs were used to track movements of about 3 million subscribers for modelling spread of the 2010 cholera outbreak. A similar study has been conducted in Pakistan to model the dynamics of Dengue (Wesolowski et al. 2015c). The Ebola disease outbreak in West Africa between 2013 and 2016 saw a collaboration between mobile companies that provided mobile phone data and researchers in trying to understand the extent to which human movement contributed to the spread of the deadly epidemic (Wesolowski et al. 2014a). Valdez et al. (2015) also investigated the impact of human movement on the Ebola epidemic by investigating the role of travel between counties in Liberia on the epidemic.

4.1 Limitations of mobility models in epidemic modelling

Several challenges and limitations remain regarding the use of mobility data and models for simulating the spatial spread of infections.

A key challenge is the relative lack of detailed movement data with which to parameterize the necessary models. Individual level data on human movement is difficult to measure and rarely reported, presenting problems in understanding individual-level variation in mobility patterns and identifying individual-level characteristics that may explain such differences. Adults and children may have very different mobility patterns (Read et al. 2014), and such patterns may be key in understanding the evolution of seasonal diseases such as influenza (Bedford et al. 2015). Despite the increasing use of CDRs for mobility studies, privacy issues generally prevent the release of individual-level information. CDR data is generally released to researchers in an aggregate form from cell phone service providers, and often cannot be used to inform movement between national borders. Daily commuting information has been widely used for epidemic modelling, but again this is often only available in an aggregated form, and does not typically provide information on non-work related movement.

A related challenge is the availability of contemporary movement data for locations experiencing an outbreak. Much of the high-resolution mobility data available is for developed countries, though increasingly this information is being made available for developing countries. During the 2013-16 West Africa Ebola epidemic, lack of good quality mobility data made modelling of transmission difficult. International airline passenger data were used to model the risk of international spread of Ebola from West Africa (Bogoch et al. 2015; Read et al. 2015), but data

capturing the local movements of people by bus or foot across the porous borders of Guinea, Liberia and Sierra Leone were not available. Local movements are thought to have played a key role in transmission of Ebola, (Cori et al. 2017).

The different model structures used by mobility studies outlined previously, highlight an important challenge for modelling mobility for epidemics, namely, what is the appropriate model structure and covariates for accurately predicting the movement of individuals? The two main model structures used, the gravity and radiation models, make different predictions of the flow of individuals between locations, which may have important consequences for estimating infection risk in particular populations (Wesolowski et al. 2015b).All applications of mobility models in epidemiology (to the best of the authors' knowledge) use a measure of population density as location-specific covariate; no studies have systematically explored a wider range of possible alternative location covariates. Many alternative covariates exist, such as environmental measures, and possible interactions between different covariates should also be considered. Many studies use geographical distance between locations as a property of both locations. However, other properties may be more appropriate, particularly for applications at small geographical scales, such as across-city movement. Journey time and travel cost may be worth exploring as potential explanatory variables, and different transport options (walking, cycling, driving or using public transport) may require different models to optimally explain mobility patterns. By extension, the movement of different types of individuals (for example, different age or socioeconomic groups) may also be best described using different models. There is a relatively low use of formal statistical testing and model selection approaches in spatial interaction modelling.

Issue of spatial scale is another important limitation to the application of gravity and radiation models in epidemic modelling. These models have been shown to not perform uniformly at different spatial scales, and it may be that different spatial scales require multiple models, instead of a single formulation. For example, when modelling within and between cities and between countries epidemic spread, the models may need further tuning to more accurately capture the underlying mobility patterns in the different settings: poorly predicting movement at one scale may render epidemic predictions inaccurate. For example, some researchers have suggested that the radiation model is most suited to predicting human mobility between larger spatial regions such as counties, and may not be applicable in predicting human mobility at finer spatial scales, such as between districts within cities, (Liang et al. 2013; Kang et al. 2015; Masucci et al. 2013; Yang et al. 2014). Yan et al. (2014) modified the standard radiation model for mobility prediction in cities using different mobility datasets such as passenger taxi commuting data in Beijing and Shenzen, mobile phone data in Abidjan, Cote d'Ivoire, and tracker

survey data in Chicago. They developed a population-weighted opportunity (PWO) radiation model which was found to predict mobility better than the standard radiation model in these different city settings. This model widens the possible destination areas of individuals, unlike the RM which assumes that people tend to choose the nearest locations with the biggest benefits. In this PWO model, the whole city encompasses the possible destination locations for individuals. Yan et al. (2014) introduced a scaling parameter to the original radiation model which takes the role of capturing the influence of the region scale and the degree of heterogeneity in the distribution of facilities that may be visited. As the bulk of social interactions are often made very close to home (Read et al. 2014), there is a need for future studies of mobility to focus on a spatial scale appropriate to the infection process.

A further challenge is the incongruity often seen between the types of movement described by mobility models and those used in epidemic simulations. Gravity models describe the flow of individuals between an origin and a destination, while epidemic models typically describe the movement of individuals. There is also a challenge in defining the quantity being modelled. Gravity models tend to focus on the flow of commuters, ignoring trajectories of individuals through space (en route to their workplace, and back); as such the intermediate locations they pass through to get to their destinations are not considered. As these other locations may also present an opportunity for infection, the spatial spread of infection may be poorly estimated. For seasonal influenza, it is thought that transmission outside the household is split evenly between the general community and school and workplaces (Ferguson et al. 2006). With sufficiently detailed data, multiple models of mobility for different purposes (work, shopping, leisure, etc) could be estimated, and epidemic prediction made using an ensemble of these different models. Additionally, models which describe the trajectory of individuals through space during a day could be developed.

Human behaviours often have strong seasonal components, such as travel during weekdays and weekends, or movements related to national holidays. Such seasonal patterns are known to be important for describing disease dynamics, and it may be important to capture these changing features within mobility models. Similarly, changes in covariate measures during the course of an epidemic may be important to incorporate into mobility models (e.g., if there are sudden changes in population density, or restrictions on movement). Additionally, travel behaviour of individuals may change in response to an epidemic (Funk et al. 2010), as was seen in Mexico during the early days of the influenza pandemic in 2009. During the SARS outbreak of 2003, there was widespread disruption on social behaviour and businesses including tourism with a drop in hotel occupancy of over 60% (Heymann

and Rodier 2004). In Hong Kong, there was a reduction in the rates of travelling (Lau et al. 2005). To adequately capture such changing epidemic features requires the development of dynamic models of mobility.

5 Conclusion

A wide range of mobility models have been developed by researchers in other disciplines, and it may benefit infectious disease modelling to consider alternative model structures and covariates. Key challenges remain, however, for the improvement of mobility prediction models and their integration within transmission models. The large amount of detailed information afforded by the widespread use of mobile technology offers an important future resource for the further development of mobility models, if privacy issues can be overcome. However, there is currently a need for greater statistical rigour regarding the construction and selection of mobility models and their covariates.

References

Anderson, James E and Eric Van Wincoop (2003). "Gravity with gravitas: a solution to the border puzzle." In: *American economic review* 93.1, pp. 170–192.

Arvis, Jean-Francois and Ben Shepherd (2013). "The Poisson quasi-maximum likelihood estimator: a solution to the 'adding up'problem in gravity models." In: *Applied Economics Letters* 20.6, pp. 515–519.

Asgari, Fereshteh, Vincent Gauthier, and Monique Becker (2013). "A survey on human mobility and its applications." In: *arXiv preprint arXiv:1307.0814*.

Balcan, Duygu, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani (2009). "Multiscale mobility networks and the spatial spreading of infectious diseases." In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21484–21489.

Barrios, José Miguel, Willem W Verstraeten, Piet Maes, Jean-Marie Aerts, Jamshid Farifteh, and Pol Coppin (2012). "Using the gravity model to estimate the spatial spread of vector-borne diseases." In: *International journal of environmental research and public health* 9.12, pp. 4346–4364.

Bedford, Trevor, Steven Riley, Ian G Barr, Shobha Broor, Mandeep Chadha, Nancy J Cox, Rodney S Daniels, C Palani Gunasekaran, Aeron C Hurt, Anne Kelso, et al. (2015). "Global circulation patterns of seasonal influenza viruses vary with antigenic drift." In: *Nature* 523.7559, p. 217.

Bengtsson, Linus, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, Kankoe Sallah, Stanislas Rebaudet, and Renaud Piarroux (2015). "Using mobile phone data to predict the spatial spread of cholera." In: *Scientific reports* 5.

Bergstrand, Jeffrey H (1985). "The gravity equation in international trade: some microeconomic foundations and empirical evidence." In: *The review of economics and statistics*, pp. 474–481.

Blumenstock, Joshua and Nathan Eagle (2010). "Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda." In: *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. ACM, p. 6.

Bogoch, Isaac I, Maria I Creatore, Martin S Cetron, John S Brownstein, Nicki Pesik, Jennifer Miniota, Theresa Tam, Wei Hu, Adriano Nicolucci, Saad Ahmed, et al. (2015). "Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak." In: *The Lancet* 385.9962, pp. 29–35.

Boni, Maciej F, Bui H Manh, Pham Q Thai, Jeremy Farrar, Tran T Hien, Nguyen T Hien, Nguyen Van Kinh, and Peter Horby (2009). "Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses." In: *BMC medicine* 7.1, p. 43.

Box, George EP and David R Cox (1964). "An analysis of transformations." In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 211–252.

Buckee, Caroline O, Amy Wesolowski, Nathan N Eagle, Elsa Hansen, and Robert W Snow (2013). "Mobile phones and malaria: modeling human and parasite travel." In: *Travel medicine and infectious disease* 11.1, pp. 15–22.

Burger, Martijn, Frank Van Oort, and Gert-Jan Linders (2009). "On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation." In: *Spatial Economic Analysis* 4.2, pp. 167–190.

Bygbjerg, IC (2012). "Double burden of noncommunicable and infectious diseases in developing countries." In: *Science* 337.6101, pp. 1499–1501.

Carrothers, Gerald AP (1956). "An historical bedew of the gravity and potential concepts of human interaction." In: *Journal of the American Institute of Planners* 22.2, pp. 94–102.

Chakraborty, A, MA Beamonte, Alan E Gelfand, MP Alonso, Pilar Gargallo, and Manuel Salvador (2013). "Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets." In: *Computational Statistics & Data Analysis* 58, pp. 292–307.

Chaney, Thomas (2018). "The gravity equation in international trade: An explanation." In: *Journal of Political Economy* 126.1, pp. 150–177.

Charaudeau, Segolene, Khashayar Pakdaman, and Pierre-Yves Boëlle (2014). "Commuter mobility and the spread of infectious diseases: application to influenza in France." In: *PloS one* 9.1, e83002.

Chen, Bi Yu, Yafei Wang, Donggen Wang, Qingquan Li, William HK Lam, and Shih-Lung Shaw (2018). "Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data." In: *Annals of the American Association of Geographers*, pp. 1–19.

Chen, Yanguang (2015). "The distance-decay function of geographical gravity model: Power law or exponential law?" In: *Chaos, Solitons & Fractals* 77, pp. 174–189.

Chen, Yanguang and Feng Xu (2013). "Redefining the attraction measure, scaling exponent and impedance function of the gravity model." In: *arXiv preprint arXiv:1306.3822*.

Chowell, Gerardo, Lisa Sattenspiel, Shweta Bansal, and Cécile Viboud (2016). "Mathematical models to characterize early epidemic growth: A review." In: *Physics of life reviews* 18, pp. 66–97.

Cliff, Andrew and Peter Haggett (2004). "Time, travel and infection." In: *British medical bulletin* 69.1, pp. 87–99.

Congdon, Peter (2010). "Random-effects models for migration attractivity and retentivity: a Bayesian methodology." In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173.4, pp. 755–774.

Cori, Anne, Christl A Donnelly, Ilaria Dorigatti, Neil M Ferguson, Christophe Fraser, Tini Garske, Thibaut Jombart, Gemma Nedjati-Gilani, Pierre Nouvellet, Steven Riley, et al. (2017). "Key data for outbreak evaluation: building on the Ebola experience." In: *Phil. Trans. R. Soc. B* 372.1721, p. 20160371.

Dalziel, Benjamin D, Babak Pourbohloul, and Stephen P Ellner (2013). "Human mobility patterns predict divergent epidemic dynamics among cities." In: *Proceedings of the Royal Society of London B: Biological Sciences* 280.1766, p. 20130763.

De Montjoye, Yves-Alexandre, Caesar A Hidalgo, Michel Verleysen, and Vincent D Blondel (2013). "Unique in the crowd: The privacy bounds of human mobility." In: *Scientific reports* 3.

Dennett, Adam (2012). *Estimating flows between geographical locations:get me started in spatial interaction modelling*. Tech. rep. Working paper 181, University College London.

Deville, Pierre, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R Stevens, Andrea E Gaughan, Vincent D Blondel, and Andrew J Tatem (2014). "Dynamic population mapping using mobile phone data." In: *Proceedings of the National Academy of Sciences* 111.45, pp. 15888–15893.

Dye, Christopher (2014). "After 2015: infectious diseases in a new era of health and development." In: *Phil. Trans. R. Soc. B* 369.1645, p. 20130426.

Eggo, Rosalind M, Simon Cauchemez, and Neil M Ferguson (2011). "Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States." In: *Journal of The Royal Society Interface* 8.55, pp. 233–243.

Ewing, GO (1974). "Gravity and linear regression models of spatial interaction: a cautionary note." In: *Economic Geography* 50.1, pp. 83–88.

Facebook (2018). URL: https://facebook.com (visited on 09/03/2018).

Ferguson, Neil M, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke (2006). "Strategies for mitigating an influenza pandemic." In: *Nature* 442.7101, p. 448.

Fik, Timothy Joseph and Gordon F Mulligan (1998). "Functional form and spatial interaction models." In: *Environment and Planning A* 30.8, pp. 1497–1507.

Finger, Flavio, Tina Genolet, Lorenzo Mari, Guillaume Constantin de Magny, Noël Magloire Manga, Andrea Rinaldo, and Enrico Bertuzzo (2016). "Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks." In: *Proceedings of the National Academy of Sciences* 113.23, pp. 6421–6426.

flickr (2018). URL: https://www.flickr.com/ (visited on 09/03/2018).

Fratianni, Michele (2007). "The gravity equation in international trade." In:

Funk, Sebastian, Marcel Salathé, and Vincent AA Jansen (2010). "Modelling the influence of human behaviour on the spread of infectious diseases: a review." In: *Journal of the Royal Society Interface* 7.50, pp. 1247–1256.

Gonzalez, Marta C, Cesar A Hidalgo, and Albert-Laszlo Barabasi (2008). "Understanding individual human mobility patterns." In: *Nature* 453.7196, pp. 779–782.

Grange, Louis de, Enrique Fernández, and Joaqun de Cea (2010). "A consolidated model of trip distribution." In: *Transportation Research Part E: Logistics and Transportation Review* 46.1, pp. 61–75.

Grietens, Koen Peeters, Charlotte Gryseels, Susan Dierickx, Melanie Bannister-Tyrrell, Suzan Trienekens, Sambunny Uk, Pisen Phoeuk, Sokha Suon, Srun Set, René Gerrets, et al. (2015). "Characterizing types of human mobility to inform differential and targeted malaria elimination strategies in Northeast Cambodia." In: *Scientific reports* 5.

Halás, Marián, Pavel Klapka, and Petr Kladivo (2014). "Distance-decay functions for daily travel-to-work flows." In: *Journal of Transport Geography* 35, pp. 107–119.

Harrington, Laura C, Thomas W Scott, Kriangkrai Lerdthusnee, Russell C Coleman, Adriana Costero, Gary G Clark, James J Jones, Sangvorn Kitthawee, Pattamaporn Kittayapong, Ratana Sithiprasasna, et al. (2005). "Dispersal of

the dengue vector Aedes aegypti within and between rural communities." In: *The American journal of tropical medicine and hygiene* 72.2, pp. 209–220.

Haynes, Kingsley E, A Stewart Fotheringham, et al. (1984). *Gravity and spatial interaction models*. Vol. 2. Sage Beverly Hills, CA.

Henry, Sabine, Paul Boyle, and Eric F Lambin (2003). "Modelling inter-provincial migration in Burkina Faso, West Africa: the role of socio-demographic and environmental factors." In: *Applied Geography* 23.2, pp. 115–136.

Heymann, David L and Guenael Rodier (2004). "SARS: lessons from a new disease." In: *Learning from SARS: Preparing for the next disease outbreak*, pp. 234–245.

Jandarov, Roman, Murali Haran, Ottar Bjørnstad, and Bryan Grenfell (2014). "Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 63.3, pp. 423–444.

Jiang, Shan, Joseph Ferreira, and Marta C González (2017). "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore." In: *IEEE Transactions on Big Data* 3.2, pp. 208–219.

Kang, Chaogui, Yu Liu, Diansheng Guo, and Kun Qin (2015). "A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint." In: *PloS one* 10.11, e0143500.

Kau, James B and CF Sirmans (1979). "The functional form of the gravity model." In: *International Regional Science Review* 4.2, pp. 127–136.

Kaufmann, Christian and Hans Briegel (2004). "Flight performance of the malaria vectors Anopheles gambiae and Anopheles atroparvus." In: *Journal of vector ecology* 29, pp. 140–153.

Krenn, Patricia J, Sylvia Titze, Pekka Oja, Andrew Jones, and David Ogilvie (2011). "Use of global positioning systems to study physical activity and the environment: a systematic review." In: *American journal of preventive medicine* 41.5, pp. 508–515.

Lau, Joseph TF, Xilin Yang, Ellie Pang, HY Tsui, Eric Wong, and Yun Kwok Wing (2005). "SARS-related perceptions in Hong Kong." In: *Emerging infectious diseases* 11.3, p. 417.

Lee, Jonggul, Bo Youl Choi, and Eunok Jung (2018). "Metapopulation model using commuting flow for national spread of the 2009 H1N1 influenza virus in the Republic of Korea." In: *Journal of Theoretical Biology* 454, pp. 320–329.

LeSage, James P and Carlos Llano (2016). "A spatial interaction model with spatially structured origin and destination effects." In: *Spatial Econometric Interaction Modelling*. Springer, pp. 171–197.

LeSage, James P and R Kelley Pace (2008). "Spatial econometric modeling of origin-destination flows." In: *Journal of Regional Science* 48.5, pp. 941–967.

Li, Xinhai, Huidong Tian, Dejian Lai, and Zhibin Zhang (2011). "Validation of the gravity model in predicting the global spread of influenza." In: *International journal of environmental research and public health* 8.8, pp. 3134–3143.

Liang, Xiao, Jichang Zhao, Li Dong, and Ke Xu (2013). "Unraveling the origin of exponential law in intra-urban human mobility." In: *Scientific reports* 3.

Lowe, Rachel, Christovam Barcellos, Caio AS Coelho, Trevor C Bailey, Giovanini Evelim Coelho, Richard Graham, Tim Jupp, Walter Massa Ramalho, Marilia Sá Carvalho, David B Stephenson, et al. (2014). "Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts." In: *The Lancet infectious diseases* 14.7, pp. 619–626.

Mabey, David, Rosanna W Peeling, Andrew Ustianowski, and Mark D Perkins (2004). "Tropical infectious diseases: diagnostics for the developing world." In: *Nature Reviews Microbiology* 2.3, pp. 231–240.

MacLachlan, Ian (2011). "Quantitative Models for Geographical Analysis Additional Required Reading for Geography 3235 (Fourth Revision)." In:

Manning, Willard G and John Mullahy (2001). "Estimating log models: to transform or not to transform?" In: *Journal of health economics* 20.4, pp. 461–494.

Marshall, John M, Sean L Wu, Samson S Kiware, Micky Ndhlovu, André Lin Ouédraogo, Mahamoudou B Touré, Hugh J Sturrock, Azra C Ghani, Neil M Ferguson, et al. (2018). "Mathematical models of human mobility of relevance to malaria transmission in Africa." In: *Scientific reports* 8.1, p. 7713.

Martınez-Zarzoso, Inmaculada (2013). "The log of gravity revisited." In: *Applied Economics* 45.3, pp. 311–327.

Masucci, A Paolo, Joan Serras, Anders Johansson, and Michael Batty (2013). "Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows." In: *Physical Review E* 88.2, p. 022812.

Meekan, Mark G, Carlos M Duarte, Juan Fernández-Gracia, Michele Thums, Ana MM Sequeira, Rob Harcourt, and Vıctor M Eguluz (2017). "The ecology of human mobility." In: *Trends in ecology & evolution* 32.3, pp. 198–210.

Neiderud, Carl-Johan (2015). "How urbanization affects the epidemiology of emerging infectious diseases." In: *Infection ecology & epidemiology* 5.1, p. 27060.

Paz-Soldan, Valerie A, Robert C Reiner Jr, Amy C Morrison, Steven T Stoddard, Uriel Kitron, Thomas W Scott, John P Elder, Eric S Halsey, Tadeusz J Kochel, Helvio Astete, et al. (2014). "Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru." In: *PLoS neglected tropical diseases* 8.6, e2888.

Pellegrini, Pasquale A and A Stewart Fotheringham (2002). "Modelling spatial choice: a review and synthesis in a migration context." In: *Progress in human geography* 26.4, pp. 487–510.

Perez-Saez, Javier, Lorenzo Mari, Enrico Bertuzzo, Renato Casagrandi, Susanne H Sokolow, Giulio A De Leo, Theophile Mande, Natalie Ceperley, Jean-Marc Froehlich, Mariam Sou, et al. (2015). "A theoretical analysis of the geography of schistosomiasis in Burkina Faso highlights the roles of human mobility and water resources development in disease transmission." In: *PLoS neglected tropical diseases* 9.10, e0004127.

Porojan, Anca (2001). "Trade flows and spatial effects: the gravity model revisited." In: *Open economies review* 12.3, pp. 265–280.

Ravenstein, Ernst Georg (1885). "The laws of migration." In: *Journal of the statistical society of London* 48.2, pp. 167–235.

Read, Jonathan M, Peter J Diggle, James Chirombo, Tom Solomon, and Matthew Baylis (2015). "Effectiveness of screening for Ebola at airports." In: *The Lancet* 385.9962, pp. 23–24.

Read, Jonathan M, Justin Lessler, Steven Riley, Shuying Wang, Li Jiu Tan, Kin On Kwok, Yi Guan, Chao Qiang Jiang, and Derek AT Cummings (2014). "Social mixing patterns in rural and urban areas of southern China." In: *Proceedings of the Royal Society of London B: Biological Sciences* 281.1785, p. 20140268.

Riley, Steven and Neil M Ferguson (2006). "Smallpox transmission and control: spatial dynamics in Great Britain." In: *Proceedings of the National Academy of Sciences* 103.33, pp. 12637–12642.

Roberts, Mick, Viggo Andreasen, Alun Lloyd, and Lorenzo Pellis (2015). "Nine challenges for deterministic epidemic models." In: *Epidemics* 10, pp. 49–53.

Ruan, Zhongyuan, Chaoqing Wang, Pak Ming Hui, and Zonghua Liu (2015). "Integrated travel network model for studying epidemics: Interplay between journeys and epidemic." In: *Scientific reports* 5.

Sarzynska, Marta, Oyita Udiani, and Na Zhang (2013). "A study of gravity-linked metapopulation models for the spatial spread of dengue fever." In: *arXiv preprint arXiv:1308.4589*.

Schicker, Rebekah Stewart, Neway Hiruy, Berhanu Melak, Woyneshet Gelaye, Belay Bezabih, Rob Stephenson, Amy E Patterson, Zerihun Tadesse, Paul M Emerson, Frank O Richards Jr, et al. (2015). "A venue-based survey of malaria, anemia and mobility patterns among migrant farm workers in Amhara region, Ethiopia." In: *PloS one* 10.11, e0143829.

Searle, Kelly M, Jailos Lubinda, Harry Hamapumbu, Timothy M Shields, Frank C Curriero, David L Smith, Philip E Thuma, and William J Moss (2017). "Characterizing and quantifying human movement patterns using GPS data

loggers in an area approaching malaria elimination in rural southern Zambia." In: *Royal Society open science* 4.5, p. 170046.

Sen, Ashish and Tony Smith (2012). *Gravity models of spatial interaction behavior*. Springer Science & Business Media.

Silal, Sheetal Prakash, Francesca Little, Karen Irma Barnes, and Lisa Jane White (2015). "Hitting a moving target: a model for malaria elimination in the presence of population movement." In: *PloS one* 10.12, e0144990.

Siliverstovs, Boriss and Dieter Schumacher (2009). "Estimating gravity equations: to log or not to log?" In: *Empirical Economics* 36.3, pp. 645–669.

Silva, JMC Santos and Silvana Tenreyro (2006). "The log of gravity." In: *The Review of Economics and statistics* 88.4, pp. 641–658.

Silva, JMC Santos and Silvana Tenreyro (2011). "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator." In: *Economics Letters* 112.2, pp. 220–222.

Simini, Filippo, Marta C González, Amos Maritan, and Albert-László Barabási (2012). "A universal model for mobility and migration patterns." In: *Nature* 484.7392, pp. 96–100.

Stoddard, Steven T, Amy C Morrison, Gonzalo M Vazquez-Prokopec, Valerie Paz Soldan, Tadeusz J Kochel, Uriel Kitron, John P Elder, and Thomas W Scott (2009). "The role of human movement in the transmission of vector-borne pathogens." In: *PLoS neglected tropical diseases* 3.7, e481.

Stouffer, Samuel A (1940). "Intervening opportunities: a theory relating mobility and distance." In: *American sociological review* 5.6, pp. 845–867.

Tatem, Andrew J, Zhuojie Huang, Clothilde Narib, Udayan Kumar, Deepika Kandula, Deepa K Pindolia, David L Smith, Justin M Cohen, Bonita Graupe, Petrina Uusiku, et al. (2014). "Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning." In: *Malaria journal* 13.1, p. 1.

Tizzoni, Michele, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza (2014). "On the use of human mobility proxies for modeling epidemics." In: *PLoS computational biology* 10.7, e1003716.

Toole, Jameson L, Yves-Alexandre de Montjoye, Marta C Gonzalez, and Alex (Sandy) Pentland (2015). "Modeling and Understanding Intrinsic Characteristics of Human Mobility." In: *Social phenomena : from data analysis to models*. Ed. by Bruno Gonçalves and Nicola Perrera. Springer. Chap. 2, pp. 15–35.

Tsutsumi, Morito and Kazuki Tamesue (2012). "Intraregional Flow Problem in Spatial Econometric Model for Origin—Destination Flows." In: *Environment and Planning B: Planning and Design* 39.6, pp. 1006–1015.

Twitter (2018). URL: https://twitter.com (visited on 09/03/2018).

UK Data Service (2018). URL: http://wicid.ukdataservice.ac.uk/ (visited on 08/11/2018).

Valdez, LD, HH Aragão Rêgo, HE Stanley, and LA Braunstein (2015). "Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies." In: *Scientific reports* 5.

Viboud, Cécile, Ottar N Bjørnstad, David L Smith, Lone Simonsen, Mark A Miller, and Bryan T Grenfell (2006). "Synchrony, waves, and spatial hierarchies in the spread of influenza." In: *science* 312.5772, pp. 447–451.

Wai, Khin Thet, Myat Phone Kyaw, Tin Oo, PeThet Zaw, Myat Htut Nyunt, Moe Thida, and Thar Tun Kyaw (2014). "Spatial distribution, work patterns, and perception towards malaria interventions among temporary mobile/migrant workers in artemisinin resistance containment zone." In: *BMC public health* 14.1, p. 1.

Weibo (2018). URL: https://weibo.com (visited on 09/03/2018).

Wesolowski, Amy, Caroline O Buckee, Linus Bengtsson, Erik Wetter, Xin Lu, and Andrew J Tatem (2014a). "Commentary: Containing the Ebola outbreak–the potential and challenge of mobile network data." In: *PLOS currents outbreaks*.

Wesolowski, Amy, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee (2012). "Quantifying the impact of human mobility on malaria." In: *Science* 338.6104, pp. 267–270.

Wesolowski, Amy, CJE Metcalf, Nathan Eagle, Janeth Kombich, Bryan T Grenfell, Ottar N Bjørnstad, Justin Lessler, Andrew J Tatem, and Caroline O Buckee (2015a). "Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data." In: *Proceedings of the National Academy of Sciences* 112.35, pp. 11114–11119.

Wesolowski, Amy, Wendy Prudhomme O'Meara, Nathan Eagle, Andrew J Tatem, and Caroline O Buckee (2015b). "Evaluating spatial interaction models for regional mobility in Sub-Saharan Africa." In: *PLOS Comput Biol* 11.7, e1004267.

Wesolowski, Amy, Taimur Qureshi, Maciej F Boni, Pål Roe Sundsøy, Michael A Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O Buckee (2015c). "Impact of human mobility on the emergence of dengue epidemics in Pakistan." In: *Proceedings of the National Academy of Sciences* 112.38, pp. 11887–11892.

Wesolowski, Amy, Gillian Stresman, Nathan Eagle, Jennifer Stevenson, Chrispin Owaga, Elizabeth Marube, Teun Bousema, Christopher Drakeley, Jonathan Cox, and Caroline O Buckee (2014b). "Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones." In: *Scientific reports* 4.

Williams, Nathalie E, Timothy A Thomas, Matthew Dunbar, Nathan Eagle, and Adrian Dobra (2015). "Measures of human mobility using mobile phone records enhanced with GIS data." In: *PloS one* 10.7, e0133630.

Wilson, Alan G (1967). "A statistical theory of spatial distribution models." In: *Transportation research* 1.3, pp. 253–269.

Xia, Yingcun, Ottar N Bjørnstad, and Bryan T Grenfell (2004). "Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics." In: *The American Naturalist* 164.2, pp. 267–281.

Yan, Xiao-Yong, Chen Zhao, Ying Fan, Zengru Di, and Wen-Xu Wang (2014). "Universal predictability of mobility patterns in cities." In: *Journal of The Royal Society Interface* 11.100, p. 20140834.

Yang, Yingxiang, Carlos Herrera, Nathan Eagle, and Marta C Gonza (2014). "Limits of Predictability in Commuting Flows in the Absence of Data for Calibration." In: *Scientific reports* 4, pp. 1–9.

Zheng, Yan-Tao, Zheng-Jun Zha, and Tat-Seng Chua (2012). "Mining travel patterns from geotagged photos." In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 3.3, p. 56.

Chapter 5

Modelling individual-level differences in human mobility in Southern China

Chirombo J, Diggle P.J., Terlouw D.J., Read J.M.

5. Modelling individual-level differences in human mobility in Southern China

Chirombo J, Diggle P.J., Terlouw D.J., Read J.M.

Abstract

Human movement plays a key role in the spread of infectious diseases. Various models have been proposed which purport to capture the average travel behaviour of individuals using underlying population density measures and are routinely used to predict movement patterns in epidemic forecast models. These models rarely distinguish between different types of individuals, nor do they incorporate individual and population level heterogeneities in travel behaviour. Here, we propose a model of movement which explicitly models an individual's multiple trip to destinations. We apply this model to data collected over the first year of a longitudinal cohort study set in and around Guangzhou Province, China.

1 Introduction

Human mobility leads to contacts which contribute to the spread of directly transmitted infectious diseases (Wesolowski et al. 2012; Viboud et al. 2006; Prem et al. 2017; Belik et al. 2011; Findlater and Bogoch 2018). Quantitatively understanding the spatio-temporal patterns of mobility is important for epidemic modelling and forecasting as well as public health control. Contacts can also help in better understanding the spread of epidemics. How far people travel from home and their rate of social interactions along their journey dictate the spread of extent of epidemics such as influenza, SARS and measles.

Several studies have been conducted to understand human social contact patterns and their relationship with infectious disease spread. For example, the polymod diary study which was carried out in Western Europe (Mossong et al. 2008). The study described social mixing, contact patterns and contact durations among participants across European countries. The study found that contact patterns displayed a strong assortativity with age. For example, school children and adults were observed to highly mix with people of the same age (Mossong et al. 2008). In another study, (Read et al. 2014) found no difference in contact rates between people in urban and rural settings in Southern China, but there were

observed differences in how far from home those contacts were made. Despite these results, it is likely that differences in mixing patterns exist between individuals. Therefore, individual level covariates can help shed more light on the underlying contact patterns.

To model human mobility, one of the common models in use is the gravity (Carrothers 1956). This model has been widely used in economics and trade. For example, to model the trade volume between two countries (Chaney 2018; Bergstrand 1985; Fratianni 2007; Anderson and Van Wincoop 2003). The gravity models tend to treat all individuals the same and do not make different models for different types of individuals (e.g. adult vs children) despite the important differences that may exist with possibly some implications for infectious disease spread.

Little work has attempted to reconcile spatial movement of individuals with social contacts and efforts to obtain information on contacts with their spatial information (e.g. distance from home or point location) have been sparse (Read et al. 2012). There is relatively little information on how contacts are distributed in space, yet the combined process of mobility and social interaction is fundamental to epidemic spread and non-pharmaceutical control of epidemics.

Here, we analyse information on journeys between pairs of locations made by participants of a cohort study in Southern China that measured social interactions and the spatial locations at which they occurred to examine the drivers of the socio-spatial process and identify key individual-level covariates that significantly affect this process.

2 Methods

2.1 Notation

We consider a set of n locations, $i : i = 1, 2, \ldots, n$, each georeferenced to the central points, x_i, of a set of grid-cells that partition the study-region. The distance between x_i and x_j is denoted by $d_{ij} = ||x_i - x_j||$, where $|| \cdot ||$ is a context-specific metric, for example straight-line distance, road distance or travel time using public transport. Here, we use straight-line distance for simplicity.

Our data consist of a sequence of observations of an individual with known home location making a social encounter in a set of locations within the larger set of n locations. This information was collected by asking individual members of the study-population where they met people. In what follows, we assume that no individual is sampled more than once or, equivalently, that multiple journeys by the same individual are independent events.

We denote sampled individuals by $k = 1, \ldots, m$, each with an associated vector of covariates a_k. We denote by b_i a vector of covariates associated with the location i. We write i_k for the home-location of individual k, and j_k for their destination, if any.

2.2 Model

We assume that journeys made between locations i and j by individual k follow independent Poisson processes with intensities λ_{ijk} per day, where

$$\log \lambda_{ijk} = a_k'\alpha + b_i'\beta + b_j'\gamma + h(d_{ij}; \phi). \tag{5.1}$$

In (5.1), $h(d; \phi)$ is a specified function of distance, with parameters ϕ; we give examples in Section 2.4.2.

We write $\theta = (\alpha, \beta, \gamma, \phi)$ for the complete set of model parameters. It follows from (5.1) and the independence assumption that the probability individual k makes no journey on the day in question is

$$q_k(\theta) = \exp\left(-\sum_{j \neq i_k} \lambda_{i_k,j,k}\right), \tag{5.2}$$

and the probability that individual k makes a journey from their home-location, i_k, to another location, j_k, is

$$p_k(\theta) = \{1 - q_k(\theta)\} \times \frac{\lambda_{i_k,j_k,k}}{\sum_{\ell \neq i_k} \lambda(i_k, \ell, k)}. \tag{5.3}$$

2.3 Likelihood ratio inference

We write the observed set of sampled journeys as $\mathcal{S} = \{(i_k, j_k) : k = 1, ..., m\}$, where $i_k = j_k$ indicates that the sampled individual made no journey on the day in question. Then, the log-likelihood for θ given $\mathcal{S}$ is

$$L(\theta) = \sum_{k=1}^{m} I(i_k, j_k) \log q_k(\theta) + \{1 - I(i_k, j_k)\} \log p_k(\theta), \tag{5.4}$$

where $I(i, j) = 1$ if $i = j$ and zero otherwise.

Maximum likelihood estimates, $\hat{\theta}$, can be obtained by maximising $L(\theta)$ with respect to θ. The large-sample variance matrix of $\hat{\theta}$ is the matrix $V(\hat{\theta}$, where $V(\theta)$ has elements

$$v_{rc} = -\frac{\partial^2 L(\theta)}{\partial \theta_r \partial \theta_c}. \tag{5.5}$$

Nested sub-models can be compared using the generalised likelihood ratio statistic, $D = 2\{L(\hat{\theta} - L(\hat{\theta}_0)\}$, where θ_0 denotes θ with p if its elements assigned fixed values. If the sub-model implied by the constrained parameter vector θ_0 is correct, the approximate sampling distribution of D is chi-squared on p degrees of freedom.

2.4 Data

We used data from the fluscape study which is a longitudinal cohort study with annual follow ups. The aim of the study was to relate risk of infection with social contact patterns in Guangdong Province, China (Jiang et al. 2016). In the survey, respondents were asked about the people that they had contacted the previous day from morning till evening. These included all the people the respondent had a face-to-face conversation or touch (Read et al. 2014). Individual level characteristics of the respondents were collected in addition to location specific covariates. The respondents were also asked about the characteristics of their contacts; such as age, total duration of the contact among other characteristics. The coordinates of the home locations of the survey respondents and the contact locations were both recorded during the survey. Data were collected from 2734 unique respondents. In this study, we used data from only 3 of the 4 surveys. Some background information of the study is described in Lessler et al. (2011)

Population density data were collected from WorldPop project at 100m resolution (*World Pop* 2018). We extracted the population densities of the home and contact locations.

2.4.1 Data processing

We processed the data by taking the following steps; Firstly, we reduced the resolution of the population density raster data from 100m x 100m by aggregating the pixels to obtain grid cells at roughly 1km x 1km resolution. During the aggregation process, the grid populations were summed to obtain higher population figures for the new grid cells. We obtained the home and contact location population densities by extracting populations using the coordinates of the two locations. In this analysis, we only considered the contact locations that were reported by participants as the potential destinations; we did not include locations for which no participant reported a contact. This is clearly a subset of potential locations in which individuals could have made a contact. The fitted models, therefore, describe the rate at which individuals made visits or return visits to a limited number of locations, and are not strictly comparable to general mobility models. To reduce the computational burden, we also restricted contacts to those within 500km of the

subjects' homes. In this study, most of the contacts were reported at an average distance of 9km from home with occasional trips to further locations. We removed 112 contact locations from the analysis that were beyond 500km of participant's home location.

Figure 5.1 shows the location of the study area within Guangdong Province, the recorded home locations, reported locations where contacts events occurred and the underlying population densities.

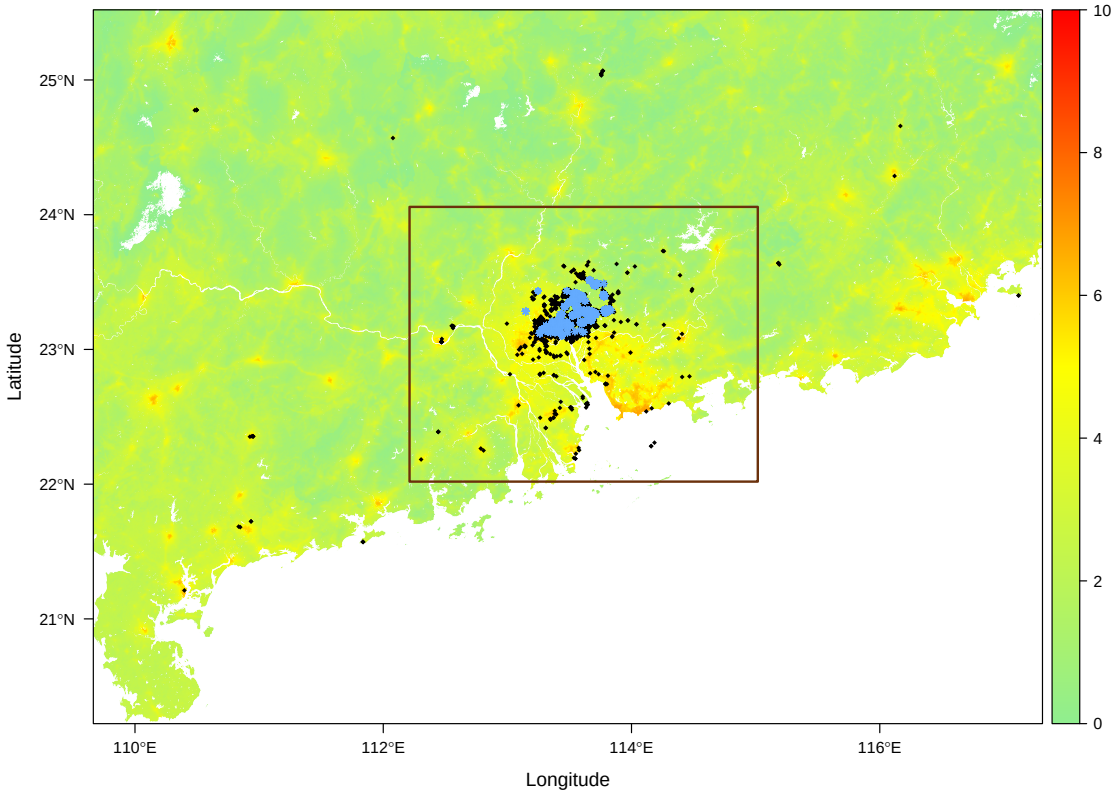

Figure 5.1: Study area within Guangdong Province in Southern China showing the home and contact locations over the study period. The blue dots are home locations while black dots represent the contact locations. Background colours green, through yellow to red show increasing population density on a log scale (density from 1 per sq. km to 22,000 per sq. km).The brown boundary shows the limit of our analysis

2.4.2 Model fitting

We fitted several models with different characteristics and different levels of complexity to aid in the model selection. Our model fitting approach was driven by our prior knowledge of the known human mobility relationships with key individual level characteristics such as age and occupation that are likely to have an effect of how people move from place to place. The first model, M1, was a gravity-like model with the population densities at origin and destination i and j and the distance as

the covariates. In the second model, M2, we fitted a similar model to M1 but now with an intercept term. Model 3 was also defined in a similar manner to M1 but with age group as a scaling effect. In our fourth model, M4, we considered another gravity-like model with age group affecting all mobility coefficients. This was done by fitting interaction terms between age groups and each of the covariates of the gravity-like model (i.e. populations densities at i and j and distance d_{ij}). The fifth model, M5, built on the gravity-like model by adding the effect of occupation category of participants as a scaling effect. This was done in order to investigate the effect of occupation on an individual's propensity to revisit contact locations. A summary of the fitted models is presented in Table 5.1.

Table 5.1: A summary of the different models fitted to the data and their description

Model	Formulation	Description
M1	$\log \lambda_{ijk} = \beta_i pop_i + \beta_j pop_j + \gamma d_{ij}$	Basic gravity like formulation, no intercept
M2	$\log \lambda_{ijk} = \alpha_k + \beta_i pop_i + \beta_j pop_j + \gamma d_{ij}$	Basic gravity like formulation with intercept term
M3	$\log \lambda_{ijk} = \alpha_k AgeGroup_k + \beta_i pop_i + \beta_j pop_j + \gamma d_{ij}$	Gravity like formulation with rate scaling by age group
M4	$\log \lambda_{ijk} = \beta_{ik} pop_i * AgeGroup_k + \beta_{jk} pop_j * AgeGroup_k + \gamma_k d_{ij} * AgeGroup_k$	Gravity formulation with age group affecting all mobility coefficients, no intercept
M5	$\log \lambda_{ijk} = \alpha_k Job_k + \beta_i pop_i + \beta_j pop_j + \gamma d_{ij}$	Gravity like formulation with scaling by occupation category

3 Results

3.1 Study demographics

During the first year of the study, 1794 enrolled of which 924(51.5%) were men and 870(48.5%) were females. In the second year, there were 2009 participants comprising 1025 (51%) males and 984 (49%) females. During the third year, a total of 1813 participants took part. Of these, 932(51.4%) were men and 878(48.4%) were females. The mean age across the genders was roughly the same (42 for females and 43 for mean). There were varied occupations for the study participants, such as full-time employment, students, part-time among others. The majority of the participants (76%) lived in rural areas around the urban centres.

In this study, we only considered individuals who made a journey from i to j to make contacts. The characteristics of this subset of individuals are shown in Table 5.2

Table 5.2: Demographic characteristics of study participants with recorded movements and contacts

	Sex $N(\%)$	
	Male	Female
Age group		
≤ 15	176(13.26)	119(9.72)
16-24	172(12.96)	120(9.80)
25-34	145(10.93)	178(14.54)
35-44	216(16.28)	236(19.28)
45-54	239(18.01)	242(19.77)
55-64	192(14.45)	174(14.22)
65-74	115(8.67)	96(7.84)
75+	64(4.82)	56(4.58)
Location		
Rural	1062(87.94)	944(77.12)
Urban	264(19.89)	279(22.79)
Emp. status		
Full-time	447(33.69)	365(29.82)
Self-employed	184(13.87)	145(11.84)
Part-time	99(7.46)	75(6.13)
Retired	167(12.58)	145(11.84)
Student	206(15.52)	144(11.76)
Home Based	9(0.69)	232(18.95)
Sick/disabled	4(0.30)	4(0.33)
Unemployed	202(51.22)	106(8.66)
Other	9(0.69)	6(0.49)
Day		
Weekday	1167(87.94)	1086(88.73)
Weekend	153(11.53)	128(10.46)

Figure 5.2 shows the differences in the distances covered by participants from their homes to destination locations. It can be observed that most of the contact events are within a relatively short distance from home at roughly around 6 km. A majority of contacts took place closer to home while a few took place far away. This pattern is similar across gender, occupation, day of the week, and location

(rural or urban). This observation of a majority of contacts occurring in the vicinity of the home locations has been highlighted in some works.

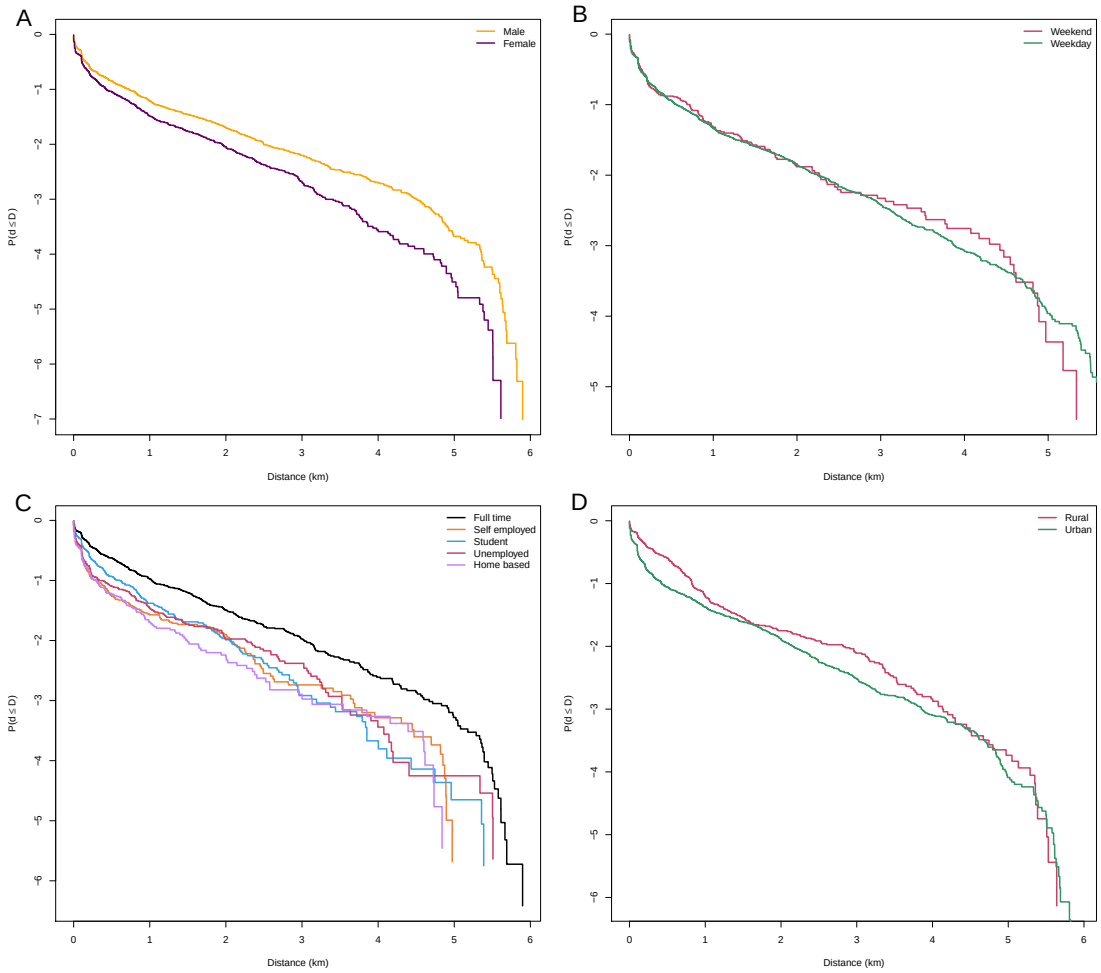

Figure 5.2: Inverse cumulative distance functions for individual movements on the log scale. The plots show the proportion of movements to make a contact at a particular distance (d) or greater from home. (A) Distance kernels for movements for males and females (B) Distance kernels for movements by day, whether weekday or weekend (C) Distance kernels for movements by employment status, and (D) Distance kernels by location, whether urban or rural

3.2 Model results

Table 5.3 is a summary of the different fitted models

Table 5.3: Parameter estimates of the fitted models

Coeff	M1	M2	M3	M5
Intercept		1.3(-2.2,4.7)		
Origin pop	-0.001(-0.002,-0.00)	-0.001(-0.02,-0.00)	-0.001(0.001,0.001)	-0.001(-0.002,-0.00)
Destn. pop	-0.004(-0.004,-0.003)	-0.003(-0.003,-0.002)	-0.003(-0.003,-0.001)	-0.003(-0.003,-0.002)
Distance	-0.005(-0.006,-0.004)	-0.004(-0.006,-0.003)	-0.005(-0.006,-0.003)	-0.006(-0.008,-0.005)
Age group				
≤ 15			0.94(0.90,1.02)	
16-24			0.95(0.90,0.99)	
25-34			0.94(0.89,0.98)	
35-44			0.74(0.70,0.77)	
45-54			0.80(0.77,0.83)	
55-64			1.13(1.09,1.16)	
65-74			0.71(0.65,0.77)	
75+			1	
Occupation				
Full-time				0.17(0.10,0.24)
Home-based				0.88(0.80,0.96)
Student				1.32(1.25,1.40)
Retired				0.70(0.62,0.78)
Self-employed				1.21(1.14,1.29)
Sick/disabled				1.00(0.70,1.31)
Part-time				0.72(0.62,0.82)
Unemployed				1
AIC	1.10	3.11	15.11	6.94

The estimates of the relationship between the populations and the distance with the age categories is given the Table 5.4

Table 5.4: Mobility model coefficients for model M4 when including an interaction effect between age group and population at origin, destination and distance

Coef	*popi* (95%)	*popj* (95%)	d_{ij} (95%)
Age group			
≤ 15	0.63(0.61,0.64)	0.31(0.30,0.33)	1.10(1.08,1.21)
16-24	0.97(0.86,0.99)	0.04(0.03,0.05)	0.59(0.51,0.66)
25-34	-0.11(-0.14,-0.09)	0.46(-0.49,5.00)	0.62(0.57,0.71)
35-44	0.33(0.23,0.54)	0.12(0.10,0.15)	0.96(0.91,1.01)
45-54	0.90(0.74,0.98)	-0.45(-0.64,-0.34)	0.005(0.003,0.007)
55-64	0.38(-1.77,2.52)	0.23(-1.92,2.38)	-0.17(-3.62,3.26)
65-74	0.53(0.51,0.54)	0.22(0.21,0.22)	0.79(0.67,0.93)
75+	0.64(0.61,0.67)	0.12(0.11,0.14)	1.66(1.46,1.86)
AIC	42.62		

In the gravity like model, M1, the rate of movement from i to j only depends on the population densities at the origin and destination locations and their distance. It is observed that the origin population has a small negative effect on the rate at

which individuals make repeated visits to locations from their home locations. As the population density at the origin increases, individuals reduce the rate at which they revisit locations. This can be interpreted as a push effect that drives people to make movements out of their home locations to make contacts elsewhere. In model M1, this push effect is relatively weak leading to fewer repeated visits. The destination density effect is also small and negative. An increase in the population density at the destination does not lead to an increase in the attractive power. This can be interpreted as a push effect. The effect of distance separating i and j is also small. All the gravity parameters are significant in the model.

In model M2, the gravity-like model with an intercept term, shares some similarities with the model M1. Model M2 again shows a negative effect for the origin, destination and distance effects. Looking at the relatively large value of the intercept term in this model, it suggests that there are other factors contributing to the rate at which repeat visits occur in addition to the population densities and the distances.

In model M3, the age group acts as a scaling effect. There is a positive effect of age on all the age groups compared to those aged above 75 years. There is also a general decreasing trend with age. The highest effect of age is observed in the 55-64 age group.

In model M4, there is an observed consistent pattern for the coefficients of age group and origin population interaction ($pop_i * AgeGroup_k$). The estimates of the age group and destination interaction ($pop_j * AgeGroup_k$) suggests that population density at potential destinations has a significant 'pull' effect. For distance, there is also a similar trend of positive interaction effects. In general, the interaction parameter estimates indicate age is an important predictor of an individual's movement rate. The parameters are shown in the Figure 5.3.

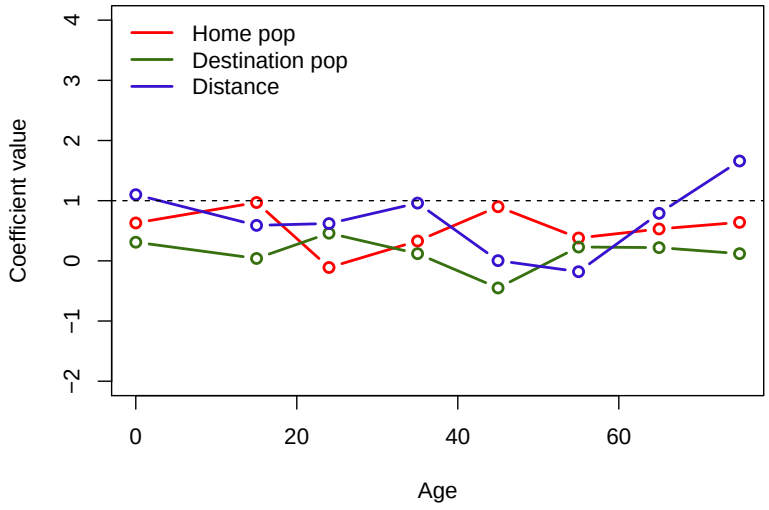

Figure 5.3: Parameter estimates of the fitted interaction model

In model M5 with scaling by the occupation, there is an observed positive effect for all occupation categories compared to the unemployed. Those on full time employment have the lowest effect. Individuals in this occupation category likely visit a single location all week and this causes their rate of returning to locations to be lower than those that visit multiple locations. Students are observed to have the highest rate of movement among all the job categories compared to the unemployed. This pattern is also observed for home-based, retired and part-time employees. Students are likely to visit more locations during the school week and also revisit them.

4 Discussion

Our modelling framework establishes a statistical framework for integrating explicit individual information to investigate mobility models which usually model mass flow of people. It allows for testing whether the inclusion of such information improves the model and whether different age groups have fundamentally different mobility models, and so behave in different patterns. For example, there is an underlying assumption that different age groups such as children and adults have the same mobility patterns leading to the use of a single model to describe the movement of both groups.

The framework also sets a foundation for the inclusion of longitudinal information for individuals making it possible to investigate whether people consistently follow a particular mobility model over a long period of time. This can be useful in the control of infectious diseases as individual's mobility patterns would be well known leading to better design of interventions targeted at a particular age group.

The effect of individual level covariates such as age, sex, and occupation provides a possible improvement to the models for aggregate flow such as the gravity model. Our study takes individual level differences in mobility patterns into account instead of modelling the total flow. In this study, we analyzed a subset of these possible models, mainly looking at age and employment as key drivers of the rate at which individuals visit known contact locations. The reduction in spatially dispersed contact in the older age groups may likely slow down contact rates and hence the reduce the risk of older people driving infections. For more effective outbreak control, this knowledge can play a key role in the design of more effective non-pharmaceutical control interventions such as movement restrictions which can be put in place to target the most likely group to transmit infection. An example of an intervention targeted at a specific groups is school closures (Cauchemez et al. 2009; House et al. 2011). It is also possible to restrict contacts in locations around communities where infections are known to be widely transmitted based on

individual qualities. Ferguson et al. (2006) found that relatively high transmission levels occur in communities around home locations where people live and not just in work locations.

Modelling individual trips between locations i and j allows for the inclusion of covariates at different scales (such as individual and location). Our approach places great emphasis on model specification, parameter estimation and model evaluation, avoiding model specification challenges commonly encountered when the gravity model is expressed as a linear model to facilitate model parameter estimation (Fik and Mulligan 1998). The formal model testing and the likelihood based inference, possible within our framework, is another strength of our approach over the common approaches that lack a well-established framework for inference. It also provides a framework for handling longitudinal data.

The main limitation of the study is that we only considered the locations where individuals reported contacts, excluding locations for which contacts were not reported. Therefore, we did not model an individual's activity space and hence it is not possible to fully characterize a person's mobility patterns. Another limitation of this study is the lack of population data for Hong Kong and Macau which are likely key destinations locations. These are big urban centres which are likely to offer some job opportunities to people from the study area and therefore attract commuters. This information was not available through WorldPop data and therefore these locations could not be considered in our models. However, travel from mainland China to these locations requires crossing a border and visa approval, so it is unlikely that these destinations are viable as daily commuting destination for the majority of our subjects and the population of Guangzhou.

For future steps, modelling the rate at which individuals make trips over an entire geographical space would be a natural extension to the approach presented in this paper. Such an approach would make it possible to cover all possible destinations for an individual instead on only a subset of locations. By looking at all possible destination locations, the model would simultaneously address both long-range and short-range travels and implications for infectious disease spread. Another future step to the model is allowing the model to handle repeated observations through the inclusion of random effects for individuals.

5 Conclusion

In this paper, we presented a model for individual journeys between location pairs, while taking into account the different individual, origin and destination characteristics that may influence the rate of making these journeys. An understanding of

these characteristics can help in further understanding the role of human movement on disease spread but also how to control the spread of infectious diseases.

References

Anderson, James E and Eric Van Wincoop (2003). "Gravity with gravitas: a solution to the border puzzle." In: *American economic review* 93.1, pp. 170–192.

Belik, Vitaly, Theo Geisel, and Dirk Brockmann (2011). "Natural human mobility patterns and spatial spread of infectious diseases." In: *Physical Review X* 1.1, p. 011001.

Bergstrand, Jeffrey H (1985). "The gravity equation in international trade: some microeconomic foundations and empirical evidence." In: *The review of economics and statistics*, pp. 474–481.

Carrothers, Gerald AP (1956). "An historical bedew of the gravity and potential concepts of human interaction." In: *Journal of the American Institute of Planners* 22.2, pp. 94–102.

Cauchemez, Simon, Neil M Ferguson, Claude Wachtel, Anders Tegnell, Guillaume Saour, Ben Duncan, and Angus Nicoll (2009). "Closure of schools during an influenza pandemic." In: *The Lancet infectious diseases* 9.8, pp. 473–481.

Chaney, Thomas (2018). "The gravity equation in international trade: An explanation." In: *Journal of Political Economy* 126.1, pp. 150–177.

Ferguson, Neil M, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke (2006). "Strategies for mitigating an influenza pandemic." In: *Nature* 442.7101, p. 448.

Fik, Timothy Joseph and Gordon F Mulligan (1998). "Functional form and spatial interaction models." In: *Environment and Planning A* 30.8, pp. 1497–1507.

Findlater, Aidan and Isaac I Bogoch (2018). "Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel." In: *Trends in parasitology*.

Fratianni, Michele (2007). "The gravity equation in international trade." In:

House, Thomas, Marc Baguelin, Albert Jan Van Hoek, Peter J White, Zia Sadique, Ken Eames, Jonathan M Read, Niel Hens, Alessia Melegaro, W John Edmunds, et al. (2011). "Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic." In: *Proceedings of the Royal Society of London B: Biological Sciences* 278.1719, pp. 2753–2760.

Jiang, Chao Qiang, Justin Lessler, Lina Kim, Kin On Kwok, Jonathan M Read, Shuying Wang, Lijiu Tan, Marisa Hast, Huachen Zhu, Yi Guan, et al. (2016). "Cohort Profile: A study of influenza immunity in the urban and rural Guangzhou region of China: the Fluscape Study." In: *International journal of epidemiology* 46.2, e16–e16.

Lessler, Justin, Derek AT Cummings, Jonathan M Read, Shuying Wang, Huachen Zhu, Gavin JD Smith, Yi Guan, Chao Qiang Jiang, and Steven Riley (2011). "Location-specific patterns of exposure to recent pre-pandemic strains of influenza A in southern China." In: *Nature communications* 2, p. 423.

Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. (2008). "Social contacts and mixing patterns relevant to the spread of infectious diseases." In: *PLoS medicine* 5.3, e74.

Prem, Kiesha, Alex R Cook, and Mark Jit (2017). "Projecting social contact matrices in 152 countries using contact surveys and demographic data." In: *PLoS computational biology* 13.9, e1005697.

Read, JM, WJ Edmunds, S Riley, J Lessler, and DAT Cummings (2012). "Close encounters of the infectious kind: methods to measure social mixing behaviour." In: *Epidemiology & Infection* 140.12, pp. 2117–2130.

Read, Jonathan M, Justin Lessler, Steven Riley, Shuying Wang, Li Jiu Tan, Kin On Kwok, Yi Guan, Chao Qiang Jiang, and Derek AT Cummings (2014). "Social mixing patterns in rural and urban areas of southern China." In: *Proceedings of the Royal Society of London B: Biological Sciences* 281.1785, p. 20140268.

Viboud, Cécile, Ottar N Bjørnstad, David L Smith, Lone Simonsen, Mark A Miller, and Bryan T Grenfell (2006). "Synchrony, waves, and spatial hierarchies in the spread of influenza." In: *science* 312.5772, pp. 447–451.

Wesolowski, Amy, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee (2012). "Quantifying the impact of human mobility on malaria." In: *Science* 338.6104, pp. 267–270.

World Pop (2018). URL: http://www.worldpop.org.uk/ (visited on 09/16/2018).

Chapter 6

Discussion

In this thesis, we investigated how different interactions between populations lead to differences in the spread of infectious diseases and reviewed and built models that considered mobility from a range of perspectives. The malaria model of Chapter 2 identified the spatial correlation of malaria cases, where human mobility is one potential mechanism generating the spatial dependence. The extension of reproduction number estimation (Chapter 3) to incorporate spatial information, generates estimated effects of one location on another: human mobility is likely to be a major factor driving these interactions. We utilized distance kernels only and population densities to measure changing disease transmission potential between a pair of locations. In Chapter 4, we reviewed different approaches to modelling human mobility from a range of disciplines, how models are used and how they are being applied in epidemiology. Finally, in Chapter 5, we considered detailed information relating where individuals make social encounters (which have the potential to spread close-contact infection) to their home locations, to identify a parsimonious model of mobility relating to contact. We explicitly modelled the individual rates of movement which are requirements for the spread of infections such as influenza, measles, and Ebola among others. We have also considered other factors, such as environmental and social characteristics, that may also lead to differences in the transmission potential of a disease, with a case study of malaria in Malawi. Studying and understanding the different processes that can drive the geographical spread of infections is an important undertaking for the control of infectious diseases.

Connectivity through mobility has been increasing, changing the face of the world in the process since the 18th century. In pre-industrial times, water ways, horses and carriages were the most common mode of transportation (Cliff et al. 1998). The industrial age brought rail transport, bicycles and steam ships which greatly increased the distance covered and improved connectivity (Van Audenhove et al. 2018). More industrialization in the 20th century brought air transport, further

expansion of roads and rail, and the coming of personal cars which have become a backbone of mobility. We are now living in an age of rapid industrialization, increasing social media usage and unprecedented digitization through computer use (Cliff et al. 1998; Van Audenhove et al. 2018). The world is becoming more urban and it has been projected that by 2050, 68% of the global population will live in urban centres, up from 55% in 2018 (UN 2018)

With increasing frequency of contacts and interaction between populations, understanding the role that human movement plays in disease spread is vital for designing, applying and evaluation of interventions. Communities that were previously isolated are becoming more accessible due to improved transport and communication links, as well as increasing urbanization. This increased rate of urbanization has been found to affect emerging infectious diseases, establishing conducive environments for new epidemics and zoonotic diseases (Neiderud 2015; Hassell et al. 2017). For malaria, as many countries, including Malawi focus on reducing transmission of malaria by 90% by the year 2030 (OrganizationWHO 2015), a key challenge will be sustaining the gains in reduction of incidence that have so far been made (Snow et al. 2017). While changing climate will affect the attainment of this goal of reducing transmission, human movement is also likely to play a key role as it helps determine the risk of introduction and reintroduction of malaria into new areas or areas where eradication has been achieved (Marshall et al. 2016; Martens and Hall 2000; Prothero 2001). In our modelling approach, the effect of human movement on malaria risk was implicitly modelled through the inclusion of random effects to capture unmeasured effects.

An understanding of human mobility and contact patterns can help to understand the likely pathways of disease spread and thus help in the design and implementation of interventions such as quarantine which is aimed at minimizing the contact between susceptible and infected individuals (Day et al. 2006). Quarantine limits the movement of people exposed to a contagious infection to see if they develop the disease. Curfews limit the movement of people in the general population. School closures are another intervention designed to reduce the effective number of contacts.

Data on human movement and mobility are therefore crucial for understanding and forecasting infectious disease spread. To achieve this goal, high resolution spatial and temporal data are needed. Individual-level mobility data are important to tease apart the mobility differences that exist between people which are critical to understanding and predicting the spread of diseases such as Ebola. The models described in this thesis capture both individual, population and location-specific information that may be responsible for the heterogeneities that exist. Therefore, the inclusion of individual and location covariates in the modelling framework is

beneficial for a better understanding of disease transmission and improved ability to apply the models to predict connectivity in populations where the measurement has not been undertaken.

Digital data sources such as mobile phone CDRs have opened the possibility of using huge datasets of finely resolved individual-level information for measuring human movement patterns (Pappalardo et al. 2015; Chen et al. 2016). There has been an increasing interest in the use of these data for disease surveillance (Hay et al. 2013; Bansal et al. 2016). Using big data characterized by fine spatial granularity provides opportunities as they increase accessibility to populations over space and time; data on personal beliefs, behaviours, and health outcomes are now available at unprecedented breadth and depth (Lee et al. 2016). Another source of digital data in addition to CDR is satellite imagery which can also provide insight into infectious disease dynamics (Lessler et al. 2016). Satellite imagery provides high spatial resolution of important drivers such as environmental, and climatic factors and population density for all locations across the globe (Sorichetta et al. 2015). These data sources are widely used in statistical models to produce maps of disease incidence, prevalence or risk (Hay and Snow 2006). Such data sources with global coverage have allowed analyses leading to an understanding of disease risk even in areas with weak disease surveillance systems. There is room for incorporating these detailed spatial data into mechanistic models that capture changes in disease risk. In the future, as digital data become more accessible in developing countries, detailed analysis to understand disease dynamics at fine scales will be possible. These data sources would further enrich the range of models that can be fitted.

However, a big challenge in using CDR data is privacy concerns (Gonzalez et al. 2008). This ethical challenge may prevent the successful utilization of the data in disease modelling studies. The highly resolved data may also lead to massive datasets which may lead to computational challenges. However, there is a chance to leverage the tools available to researchers such as high-performance computing to analyze these data and provide insights into disease transmission. With CDRs, it is not possible to measure human movement at a resolution finer than the cell tower resolution (Wesolowski et al. 2016).

Another challenge is the crude nature of the data and the lack of individual-level characteristics necessary to model the individual level differences in mobility. The CDR, which is very detailed data source, are usually aggregated due to privacy concerns. Though it is technically possible to obtain individual-level trajectories with CDR data, it is not usually done in practice. Studies such as Wesolowski et al. (2012) used aggregate data to quantify the impact of human mobility on malaria transmission. In situations where individual-level data have been successfully collected through surveys, for example, other concerns such as representativeness

arise. Robust study designs and novel data collection methods are needed to capture detailed data on mobility such as contact location and duration. In resource-limited settings, it may be challenging to design such mobility studies at a large scale leading to low sample sizes. For example, the use of digital trackers, though excellent in capturing individual travel patterns, may be expensive to implement at a large scale. The resulting data may therefore not fully represent the underlying population. Use of digital data collection methods is also prone to technological fail and wrong use of the gadgets thus leading to unreliable results. For example, the use of GPS trackers may not work well in obstructed environments such as inside houses. The data also generally lacks repeated measurements of individuals. This makes it difficult to track individual-level changes in mobility behaviour over time.

In addition to the ethical and technological challenges, the CDR data also needs validation. In some cases, it has been shown that travel distances tend to be lower when measured by CDR (Zhao et al. 2016). The general differences in ownership of mobile phones among different groups that may have different mobility rates make a generalization to the wider population a challenge (Wesolowski et al. 2013). The assumption that CDRs present a representative sample may not hold in some settings due to some known biases that exist such as higher mobile phone ownership among wealthier urban males (Wesolowski et al. 2014b). Therefore, there is a need to validate the CDR data using additional data on socio-economic status, mobile phone ownership, and usage to determine their representativeness (Wesolowski et al. 2013). Comparatively, mobility data from surveys have the advantage of being more representative as the design of the data collection process is under the control of the researchers. The surveys also make it possible to collect individual-level information leading to deeper understanding of the population under study. Therefore, combining CDR and survey data provides the best of both data approaches.

Another challenge in dealing with epidemic data is that there is a lack of a standardized data reporting format which may negatively affect analyses to evaluate and guide the application of effective interventions (Cori et al. 2017). Consequently, some important data may not be collected due to lack of resources and adequate time. This lack of adequate data presents a challenge in epidemic modelling. Finnie et al. (2016) have developed a standard for the transmission and storage of epidemiological data.

To facilitate and promote the use of mobility data in epidemiology, there is a need to make the data more accessible particularly for researchers in developing countries where there is the most potential for infectious disease health applications. One way of achieving this is by having a database for human movement linked to disease outcomes. For a start, this mobility data can be derived from censuses and

large-scale surveys such as DHS and then linked with existing disease surveillance systems such as DHIS, a web-based database for routine data. For example, climate data has been integrated with the DHIS in Tanzania as a way of facilitating the use of climate information in the health sector (Thomson 2018). A similar initiative can be done to integrate the hospital-based data with mobility data can help public health officials in disease surveillance. By making the mobility data freely accessible, its application in epidemiology would likely increase.

At present, it is also challenging to obtain mobility data for an extended period due to costs and complexities in analyzing the data among other reasons. For example, CDR data are typically released for a period ranging from months to a few years. With very short time series, it may be impossible to investigate seasonality patterns in people's mobility patterns. As a result, it is difficult to assess the presence of inter-annual variability in mobility patterns which may be of interest. Ideally, mobility data should be for a relatively long period of time to permit investigation seasonality at different temporal scales. When coupled with long time series of epidemiological data, it may lead to the establishment of an early warning system for diseases such as malaria and dengue which takes human movement into consideration

For successful studies on human mobility and its impact on disease spread, it is important to engage the various stakeholders such as mobile phone operators early on to raise awareness to the potential use of the CDR data in epidemiology. Ideally, this should be done way before a study to obtain buy-in. In most settings, human mobility studies using CDR have not yet been done and local mobile phone operators have not been engaged, leading to an information gap. If the companies are aware of the potential use of their data, they may be receptive when approached to provide data for a project.

For future research, it is important to focus on individual level differences on mobility patterns. Infectious diseases are spread when a susceptible person comes into contact with an infectious individual. Therefore, an understanding of mobility patterns at the individual level can help better understand the transmission of the disease at the population level. More research into statistical and mathematical models that use individual-level data is therefore needed. Research should also focus on study and sampling designs for mobility studies which are most cost-beneficial to allow widespread adoption in resource-limited settings. This will allow for the use of modern data collection methods even in low resource settings.

There is also a need for other techniques to understand the hidden patterns in huge datasets such as mobility data from CDRs. Machine learning algorithms provide a way to understand these massive data for understanding human movement patterns at different spatial and temporal scales and provide a mechanism for making

predictions based on the observed data (Toch et al. 2018). Fortunately, the power of machine learning increases with an increase in the size of the data. Therefore, massive datasets present an opportunity rather than a challenge. For human mobility studies, both supervised and unsupervised machine learning approaches can be applied. In unsupervised machine learning, the focus lies on finding the hidden structure and can therefore play a key role in understanding patterns of mobility in huge datasets. On the other hand, supervised machine learning is concerned with making predictions (Kotsiantis et al. 2007; Klassen et al. 2018). Both approaches are important, and a successful application of machine learning can, therefore, help bring out complex mobility patterns useful for predicting disease spread. In addition to machine learning, there is a need for widespread use of simulation studies at different scales which can then be validated by available datasets.

Data visualization for human movement is another area that must be focused on in the future. Advanced data visualization techniques need to be developed for complex mobility data for the purpose of communicating with a wider audience. The use of common static visualization tools may not best capture the complexities of human movement behaviour. For this reason, there is a need to develop web-based interactive visualization tools that can accept different mobility parameters and produce different mobility patterns based on different scenarios.

There is also a need for further work on understanding human mobility patterns in times of disasters such as flooding, drought, and wars which cause widespread displacement both within and between countries (Song et al. 2017). This can be achieved by investigating different scenarios in the modelling framework. This understanding can help understand the negative impacts both within and between countries to help authorities plan for interventions. For example, this knowledge may help authorities in resource allocation and anticipate the introduction of possible disease outbreaks. As data during these events may be difficult to obtain, prior knowledge of the effects of disasters on human mobility has the potential to improve governments' response. In many parts of the world, displacements are becoming more widespread due to conflict and natural disasters such as flooding caused by climate change.

More work is also needed on mobility modelling to capture a comprehensive pattern for an individual and relax assumptions that the person's current location is independent of past locations. A thorough mobility approach needs to capture this inherent dependency in the mobility structure. Models are also needed that model the entire space that a person may visit and not just discrete set of visited locations. In this way, the resulting model would capture an entire activity space for an individual. This approach is likely to provide a detailed insight into human

mobility patterns.

Conclusion

In conclusion, the thesis makes the following contributions: (1) a modelling framework for integrating climate and non-climate drivers into the analyses of routine malaria data at the facility level and mapping the estimates for easy to use tools for malaria control programmes. Depending on data availability, the model can also explicitly incoporate mobility information; (2) a methodology for calculating localized spatial-temporal reproductive numbers accounting for spatial interaction effects; (3) provides a detailed historical account of spatial interaction models and presents an overview of their current use in epidemic modelling, as well as future directions for research; (4) a modelling framework for individual differences in movement while taking into account the different personal and location-specific characteristics that are likely to drive observed differences in movement rates.

When ethical and other barriers are overcome in the future, it is likely to lead to improved detail and accessibility of data and our understanding of mobility patterns and the application to epidemics is likely to increase leading to improved control.

Bibliography

(NMCP)[Malawi], National Malaria Control Programme and ICF International (2012). "Malaria Indicator Survey (MIS) 2012." In:

Abeku, Tarekegn Abose, Gerrit J van Oortmarssen, Gerard Borsboom, Sake J de Vlas, and JDF Habbema (2003). "Spatial and temporal variations of malaria epidemic risk in Ethiopia: factors involved and implications." In: *Acta tropica* 87.3, pp. 331–340.

Alegana, Victor A, Peter M Atkinson, Jim A Wright, Richard Kamwi, Petrina Uusiku, Stark Katokele, Robert W Snow, and Abdisalan M Noor (2013). "Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial–temporal models." In: *Spatial and spatio-temporal epidemiology* 7, pp. 25–36.

Alirol, Emilie, Laurent Getaz, Beat Stoll, François Chappuis, and Louis Loutan (2011). "Urbanisation and infectious diseases in a globalised world." In: *The Lancet infectious diseases* 11.2, pp. 131–141.

Altizer, Sonia, Richard S Ostfeld, Pieter TJ Johnson, Susan Kutz, and C Drew Harvell (2013). "Climate change and infectious diseases: from evidence to a predictive framework." In: *science* 341.6145, pp. 514–519.

Amexo, Mark, Rachel Tolhurst, Guy Barnish, and Imelda Bates (2004). "Malaria misdiagnosis: effects on the poor and vulnerable." In: *The Lancet* 364.9448, pp. 1896–1898.

Amouzou, Agbessi, Willie Kachaka, Benjamin Banda, Martina Chimzimu, Kenneth Hill, and Jennifer Bryce (2013). "Monitoring child survival in 'real-time' using routine health facility records: results from Malawi." In: *Tropical Medicine & International Health* 18.10, pp. 1231–1239.

Anderson, Craig, Duncan Lee, and Nema Dean (2014). "Identifying clusters in Bayesian disease mapping." In: *Biostatistics*, kxu005.

Anderson, James E (2010). *The gravity model*. Tech. rep. National Bureau of Economic Research.

Anderson, James E and Eric Van Wincoop (2003). "Gravity with gravitas: a solution to the border puzzle." In: *American economic review* 93.1, pp. 170–192.

Arvis, Jean-Francois and Ben Shepherd (2013). "The Poisson quasi-maximum likelihood estimator: a solution to the 'adding up'problem in gravity models." In: *Applied Economics Letters* 20.6, pp. 515–519.

Asgari, Fereshteh, Vincent Gauthier, and Monique Becker (2013). "A survey on human mobility and its applications." In: *arXiv preprint arXiv:1307.0814*.

Balcan, Duygu, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani (2009). "Multiscale mobility networks and the spatial spreading of infectious diseases." In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21484–21489.

Bansal, Shweta, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, and Cécile Viboud (2016). "Big data for infectious disease surveillance and modeling." In: *The Journal of infectious diseases* 214.suppl_4, S375–S379.

Barrios, José Miguel, Willem W Verstraeten, Piet Maes, Jean-Marie Aerts, Jamshid Farifteh, and Pol Coppin (2012). "Using the gravity model to estimate the spatial spread of vector-borne diseases." In: *International journal of environmental research and public health* 9.12, pp. 4346–4364.

Bauke, Heiko (2007). "Parameter estimation for power-law distributions by maximum likelihood methods." In: *European Physical Journal B–Condensed Matter* 58.2.

Beck-Johnson, Lindsay M, William A Nelson, Krijn P Paaijmans, Andrew F Read, Matthew B Thomas, and Ottar N Bjørnstad (2017). "The importance of temperature fluctuations in understanding mosquito population dynamics and malaria risk." In: *Open Science* 4.3, p. 160969.

Bedford, Trevor, Steven Riley, Ian G Barr, Shobha Broor, Mandeep Chadha, Nancy J Cox, Rodney S Daniels, C Palani Gunasekaran, Aeron C Hurt, Anne Kelso, et al. (2015). "Global circulation patterns of seasonal influenza viruses vary with antigenic drift." In: *Nature* 523.7559, p. 217.

Beiró, Mariano G, André Panisson, Michele Tizzoni, and Ciro Cattuto (2016). "Predicting human mobility through the assimilation of social media traces into mobility models." In: *EPJ Data Science* 5.1, p. 30.

Belik, Vitaly, Theo Geisel, and Dirk Brockmann (2011). "Natural human mobility patterns and spatial spread of infectious diseases." In: *Physical Review X* 1.1, p. 011001.

Bengtsson, Linus, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, Kankoe Sallah, Stanislas Rebaudet, and Renaud Piarroux (2015a). "Using mobile phone data to predict the spatial spread of cholera." In: *Scientific reports* 5.

Bengtsson, Linus, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, Kankoe Sallah, Stanislas Rebaudet, and Renaud Piarroux (2015b). "Using mobile phone data to predict the spatial spread of cholera." In: *Scientific reports* 5, p. 8923.

Bengtsson, Linus, Xin Lu, Anna Thorson, Richard Garfield, and Johan Von Schreeb (2011). "Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti." In: *PLoS medicine* 8.8, e1001083.

Bergstrand, Jeffrey H (1985). "The gravity equation in international trade: some microeconomic foundations and empirical evidence." In: *The review of economics and statistics*, pp. 474–481.

Bernardinelli, L, D Clayton, C Pascutto, C Montomoli, M Ghislandi, and M Songini (1995). "Bayesian analysis of space-time variation in disease risk." In: *Statistics in medicine* 14.21-22, pp. 2433–2443.

Besag, Julian, Jeremy York, and Annie Mollié (1991). "Bayesian image restoration, with two applications in spatial statistics." In: *Annals of the institute of statistical mathematics* 43.1, pp. 1–20.

Best, Nicky, Sylvia Richardson, and Andrew Thomson (2005). "A comparison of Bayesian spatial models for disease mapping." In: *Statistical methods in medical research* 14.1, pp. 35–59.

Bettencourt, Luis MA and Ruy M Ribeiro (2008). "Real time bayesian estimation of the epidemic potential of emerging infectious diseases." In: *PLoS One* 3.5, e2185.

Blanford, Justine I, Zhuojie Huang, Alexander Savelyev, and Alan M MacEachren (2015). "Geo-located tweets. enhancing mobility maps and capturing cross-border movement." In: *PloS one* 10.6, e0129202.

Blumenstock, Joshua and Nathan Eagle (2010). "Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda." In: *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. ACM, p. 6.

Bødker, Ren, William Kisinza, Robert Malima, Hamisi Msangeni, and Steve Lindsay (2000). "Resurgence of malaria in the Usambara mountains, Tanzania, an epidemic of drug-resistant parasites." In: *Global Change and Human Health* 1.2, pp. 134–153.

Bogoch, Isaac I, Oliver J Brady, Moritz UG Kraemer, Matthew German, Marisa I Creatore, Manisha A Kulkarni, John S Brownstein, Sumiko R Mekaru, Simon I Hay, Emily Groot, et al. (2016). "Anticipating the international spread of Zika virus from Brazil." In: *The Lancet* 387.10016, pp. 335–336.

Bogoch, Isaac I, Maria I Creatore, Martin S Cetron, John S Brownstein, Nicki Pesik, Jennifer Miniota, Theresa Tam, Wei Hu, Adriano Nicolucci, Saad Ahmed, et al. (2015). "Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak." In: *The Lancet* 385.9962, pp. 29–35.

Boni, Maciej F, Bui H Manh, Pham Q Thai, Jeremy Farrar, Tran T Hien, Nguyen T Hien, Nguyen Van Kinh, and Peter Horby (2009). "Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses." In: *BMC medicine* 7.1, p. 43.

Box, George EP and David R Cox (1964). "An analysis of transformations." In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 211–252.

Boyd, Heather A, W Dana Flanders, David G Addiss, and Lance A Waller (2005). "Residual spatial correlation between geographically referenced observations: a Bayesian hierarchical modeling approach." In: *Epidemiology*, pp. 532–541.

Breslow, N. E. and D. G. Clayton (1993). "Approximate Inference in Generalized Linear Mixed Models." In: *Journal of the American Statistical Association* 88.421, pp. 9–25.

Brix, Anders and Peter J Diggle (2001). "Spatiotemporal prediction for log-Gaussian Cox processes." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.4, pp. 823–841.

Brockmann, Dirk, Lars Hufnagel, and Theo Geisel (2006). "The scaling laws of human travel." In: *Nature* 439.7075, pp. 462–465.

Buckee, Caroline O, Amy Wesolowski, Nathan N Eagle, Elsa Hansen, and Robert W Snow (2013). "Mobile phones and malaria: modeling human and parasite travel." In: *Travel medicine and infectious disease* 11.1, pp. 15–22.

Burger, Martijn, Frank Van Oort, and Gert-Jan Linders (2009). "On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation." In: *Spatial Economic Analysis* 4.2, pp. 167–190.

Bürkner, Paul-Christian (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan." In: *Journal of Statistical Software* 80.1, pp. 1–28. DOI: 10.18637/jss.v080.i01.

Bygbjerg, IC (2012). "Double burden of noncommunicable and infectious diseases in developing countries." In: *Science* 337.6101, pp. 1499–1501.

Caminade, Cyril, Sari Kovats, Joacim Rocklov, Adrian M Tompkins, Andrew P Morse, Felipe J Colón-González, Hans Stenlund, Pim Martens, and Simon J Lloyd (2014). "Impact of climate change on global malaria distribution." In: *Proceedings of the National Academy of Sciences* 111.9, pp. 3286–3291.

Campbell-Lendrum, Diarmid, Lucien Manga, Magaran Bagayoko, and Johannes Sommerfeld (2015). "Climate change and vector-borne diseases: what are the implications for public health research and policy?" In: *Phil. Trans. R. Soc. B* 370.1665, p. 20130552.

Carrothers, Gerald AP (1956). "An historical bedew of the gravity and potential concepts of human interaction." In: *Journal of the American Institute of Planners* 22.2, pp. 94–102.

Cash, BA, X Rodó, J Ballester, MJ Bouma, A Baeza, R Dhiman, and M Pascual (2013). "Malaria epidemics and the influence of the tropical South Atlantic on the Indian monsoon." In: *Nature climate change* 3.5, pp. 502–507.

Castro, Marcia C, Shogo Kanamori, Khadija Kannady, Sigsbert Mkude, Gerry F Killeen, and Ulrike Fillinger (2010). "The importance of drains for the larval development of lymphatic filariasis and malaria vectors in Dar es Salaam, United Republic of Tanzania." In: *PLoS Negl Trop Dis* 4.5, e693.

Cauchemez, Simon, Pierre-Yves Boëlle, Christl A Donnelly, Neil M Ferguson, Guy Thomas, Gabriel M Leung, Anthony J Hedley, Roy M Anderson, and Alain-Jacques Valleron (2006a). "Real-time estimates in early detection of SARS." In: *Emerging infectious diseases* 12.1, p. 110.

Cauchemez, Simon, Pierre-Yves Boëlle, Christl A Donnelly, Neil M Ferguson, Guy Thomas, Gabriel M Leung, Anthony J Hedley, Roy M Anderson, and Alain-Jacques Valleron (2006b). "Real-time estimates in early detection of SARS." In: *Emerging infectious diseases* 12.1, p. 110.

Cauchemez, Simon, Pierre-Yves Boëlle, Guy Thomas, and Alain-Jacques Valleron (2006c). "Estimating in real time the efficacy of measures to control emerging communicable diseases." In: *American journal of epidemiology* 164.6, pp. 591–597.

Cauchemez, Simon, Neil M Ferguson, Claude Wachtel, Anders Tegnell, Guillaume Saour, Ben Duncan, and Angus Nicoll (2009). "Closure of schools during an influenza pandemic." In: *The Lancet infectious diseases* 9.8, pp. 473–481.

Chaki, Prosper P, Nicodem J Govella, Bryson Shoo, Abdullah Hemed, Marcel Tanner, Ulrike Fillinger, and Gerry F Killeen (2009). "Achieving high coverage of larval-stage mosquito surveillance: challenges for a community-based mosquito control programme in urban Dar es Salaam, Tanzania." In: *Malaria Journal* 8.1, p. 1.

Chakraborty, A, MA Beamonte, Alan E Gelfand, MP Alonso, Pilar Gargallo, and Manuel Salvador (2013). "Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets." In: *Computational Statistics & Data Analysis* 58, pp. 292–307.

Chaney, Thomas (2018). "The gravity equation in international trade: An explanation." In: *Journal of Political Economy* 126.1, pp. 150–177.

Charaudeau, Segolene, Khashayar Pakdaman, and Pierre-Yves Boëlle (2014). "Commuter mobility and the spread of infectious diseases: application to influenza in France." In: *PloS one* 9.1, e83002.

Chaulagai, Chet N, Christon M Moyo, Jaap Koot, Humphrey BM Moyo, Thokozani C Sambakunsi, Ferdinand M Khunga, and Patrick D Naphini (2005). "Design

and implementation of a health management information system in Malawi: issues, innovations and results." In: *Health policy and planning* 20.6, pp. 375–384.

Chen, Bi Yu, Yafei Wang, Donggen Wang, Qingquan Li, William HK Lam, and Shih-Lung Shaw (2018). "Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data." In: *Annals of the American Association of Geographers*, pp. 1–19.

Chen, Cynthia, Jingtao Ma, Yusak Susilo, Yu Liu, and Menglin Wang (2016). "The promises of big data and small data for travel behavior (aka human mobility) analysis." In: *Transportation research part C: emerging technologies* 68, pp. 285–299.

Chen, Yanguang (2015). "The distance-decay function of geographical gravity model: Power law or exponential law?" In: *Chaos, Solitons & Fractals* 77, pp. 174–189.

Chen, Yanguang and Feng Xu (2013). "Redefining the attraction measure, scaling exponent and impedance function of the gravity model." In: *arXiv preprint arXiv:1306.3822*.

CHIRPS (2018). URL: http://iridl.ldeo.columbia.edu/SOURCES/.UCSB/.CHIRPS/.v2p0/.monthly/.global/.precipitation (visited on 09/03/2018).

Chowell, Gerardo, Lisa Sattenspiel, Shweta Bansal, and Cécile Viboud (2016). "Mathematical models to characterize early epidemic growth: A review." In: *Physics of life reviews* 18, pp. 66–97.

Christian, Kira A, Kashef Ijaz, Scott F Dowell, Catherine C Chow, Rohit A Chitale, Joseph S Bresee, Eric Mintz, Mark A Pallansch, Steven Wassilak, Eugene McCray, et al. (2013). "What we are watching—five top global infectious disease threats, 2012: a perspective from CDC's Global Disease Detection Operations Center." In: *Emerging health threats journal* 6.1, p. 20632.

Chun, Byung Chul (2016). "Understanding and modeling the super-spreading events of the Middle East respiratory syndrome outbreak in Korea." In: *Infection & chemotherapy* 48.2, pp. 147–149.

Churcher, Thomas S, Jean-François Trape, and Anna Cohuet (2015). "Human-to-mosquito transmission efficiency increases as malaria is controlled." In: *Nature communications* 6.

Cinnamon, Jonathan, Sarah K Jones, and W Neil Adger (2016). "Evidence and future potential of mobile phone data for disease disaster management." In: *Geoforum* 75, pp. 253–264.

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman (2009). "Power-law distributions in empirical data." In: *SIAM review* 51.4, pp. 661–703.

Cleveland, Robert B, William S Cleveland, Jean E McRae, and Irma Terpenning (1990). "STL: A seasonal-trend decomposition procedure based on loess." In: *Journal of Official Statistics* 6.1, pp. 3–73.

Cleveland, William S (1979). "Robust locally weighted regression and smoothing scatterplots." In: *Journal of the American statistical association* 74.368, pp. 829–836.

Cliff, Andrew and Peter Haggett (2004). "Time, travel and infection." In: *British medical bulletin* 69.1, pp. 87–99.

Cliff, Andrew, Peter Haggett, Matthew Smallman-Raynor, et al. (1998). *Deciphering global epidemics: analytical approaches to the disease records of world cities, 1888-1912*. Vol. 26. Cambridge University Press.

Cochrane, RA (1975). "A possible economic basis for the gravity model." In: *Journal of Transport Economics and Policy*, pp. 34–49.

Congdon, Peter (2010). "Random-effects models for migration attractivity and retentivity: a Bayesian methodology." In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173.4, pp. 755–774.

Connor, Stephen J and Gilma C Mantilla (2008). "Integration of seasonal forecasts into early warning systems for climate-sensitive diseases such as malaria and dengue." In: *Seasonal Forecasts, Climatic Change and Human Health*. Springer, pp. 71–84.

Cori, Anne, Christl A Donnelly, Ilaria Dorigatti, Neil M Ferguson, Christophe Fraser, Tini Garske, Thibaut Jombart, Gemma Nedjati-Gilani, Pierre Nouvellet, Steven Riley, et al. (2017). "Key data for outbreak evaluation: building on the Ebola experience." In: *Phil. Trans. R. Soc. B* 372.1721, p. 20160371.

Cori, Anne, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez (2013). "A new framework and software to estimate time-varying reproduction numbers during epidemics." In: *American journal of epidemiology* 178.9, pp. 1505–1512.

Cosner, C, JC Beier, RS Cantrell, D Impoinvil, L Kapitanski, MD Potts, A Troyo, and S Ruan (2009). "The effects of human movement on the persistence of vector-borne diseases." In: *Journal of theoretical biology* 258.4, pp. 550–560.

Cowling, Benjamin J, Vicky J Fang, Steven Riley, JS Malik Peiris, and Gabriel M Leung (2009). "Estimation of the serial interval of influenza." In: *Epidemiology (Cambridge, Mass.)* 20.3, p. 344.

Cox, David R (1972). "Regression models and life-tables." In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187–220.

Cox, David R (1975). "Partial likelihood." In: *Biometrika* 62.2, pp. 269–276.

Curry, Leslie (1972). "A spatial analysis of gravity flows." In: *Regional Studies* 6.2, pp. 131–147.

Dalziel, Benjamin D, Babak Pourbohloul, and Stephen P Ellner (2013). "Human mobility patterns predict divergent epidemic dynamics among cities." In: *Proceedings of the Royal Society of London B: Biological Sciences* 280.1766, p. 20130763.

Day, Troy, Andrew Park, Neal Madras, Abba Gumel, and Jianhong Wu (2006). "When is quarantine a useful control strategy for emerging infectious diseases?" In: *American Journal of Epidemiology* 163.5, pp. 479–485.

De Montjoye, Yves-Alexandre, Caesar A Hidalgo, Michel Verleysen, and Vincent D Blondel (2013). "Unique in the crowd: The privacy bounds of human mobility." In: *Scientific reports* 3.

De Silva, Prathiba M and John M Marshall (2012). "Factors contributing to urban malaria transmission in sub-Saharan Africa: a systematic review." In: *Journal of tropical medicine* 2012.

Dennett, Adam (2012). *Estimating flows between geographical locations:get me started in spatial interaction modelling*. Tech. rep. Working paper 181, University College London.

Deville, Pierre, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R Stevens, Andrea E Gaughan, Vincent D Blondel, and Andrew J Tatem (2014). "Dynamic population mapping using mobile phone data." In: *Proceedings of the National Academy of Sciences* 111.45, pp. 15888–15893.

Deville, Pierre, Chaoming Song, Nathan Eagle, Vincent D Blondel, Albert-László Barabási, and Dashun Wang (2016). "Scaling identity connects human mobility and social interactions." In: *Proceedings of the National Academy of Sciences* 113.26, pp. 7047–7052.

DHIS (2018). URL: dhis2.health.gov.mw/ (visited on 09/03/2018).

Dietz, Klaus (1993). "The estimation of the basic reproduction number for infectious diseases." In: *Statistical methods in medical research* 2.1, pp. 23–41.

Diggle, Peter (1985). "A Kernel Method for Smoothing Point Process Data." In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 34.2, pp. 138–147.

Diggle, Peter J (2006a). "Spatio-temporal point processes, partial likelihood, foot and mouth disease." In: *Statistical methods in medical research* 15.4, pp. 325–336.

Diggle, Peter J (2006b). "Spatio-temporal point processes, partial likelihood, foot and mouth disease." In: *Statistical methods in medical research* 15.4, pp. 325–336.

Diggle, Peter J and Emanuele Giorgi (2015). "Model-Based Geostatistics for Prevalence Mapping in Low-Resource Settings." In: *Journal of the American Statistical Association* just-accepted, pp. 1–42.

Diggle, Peter J, Yongtao Guan, Anthony C Hart, Fauzia Paize, and Michelle Stanton (2012). "Estimating individual-level risk in spatial epidemiology using spatially aggregated information on the population at risk." In: *Journal of the American Statistical Association*.

Diggle, Peter J, Irene Kaimi, and Rosa Abellana (2010). "Partial-Likelihood Analysis of Spatio-Temporal Point-Process Data." In: *Biometrics* 66.2, pp. 347–354.

Diggle, Peter J, Paula Moraga, Barry Rowlingson, Benjamin M Taylor, et al. (2013). "Spatial and spatio-temporal Log-Gaussian Cox processes: extending the geostatistical paradigm." In: *Statistical Science* 28.4, pp. 542–563.

Diggle, Peter J, JA Tawn, and RA Moyeed (1998). "Model-based geostatistics." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47.3, pp. 299–350.

Diggle, Peter J, Madeleine C Thomson, OF Christensen, B Rowlingson, Valérie Obsomer, Jacques Gardon, S Wanji, I Takougang, P Enyong, J Kamgno, et al. (2007). "Spatial modelling and the prediction of Loa loa risk: decision making under uncertainty." In: *Annals of Tropical Medicine & Parasitology* 101.6, pp. 499–509.

Diggle, Peter, Rana Moyeed, Barry Rowlingson, and Madeleine Thomson (2002). "Childhood malaria in the Gambia: a case-study in model-based geostatistics." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51.4, pp. 493–506.

Diggle, Peter, Barry Rowlingson, and Ting-li Su (2005). "Point process methodology for on-line spatio-temporal disease surveillance." In: *Environmetrics* 16.5, pp. 423–434.

Dinku, T, R Cousin, J Corral, and A Vadillo (2016). "The ENACTS Approach: Transforming climate services in Africa one country at a time." In: *World Policy Papers*, pp. 1–24.

Dye, Christopher (2014). "After 2015: infectious diseases in a new era of health and development." In: *Phil. Trans. R. Soc. B* 369.1645, p. 20130426.

Eames, Ken TD and Matt J Keeling (2002). "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases." In: *Proceedings of the National Academy of Sciences* 99.20, pp. 13330–13335.

Eggo, Rosalind M, Simon Cauchemez, and Neil M Ferguson (2010). "Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States." In: *Journal of The Royal Society Interface*, rsif20100216.

Eggo, Rosalind M, Simon Cauchemez, and Neil M Ferguson (2011). "Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States." In: *Journal of The Royal Society Interface* 8.55, pp. 233–243.

Eubank, Stephen (2005). "Network based models of infectious disease spread." In: *Japanese journal of infectious diseases* 58.6, S.

Ewing, GO (1974). "Gravity and linear regression models of spatial interaction: a cautionary note." In: *Economic Geography* 50.1, pp. 83–88.

Facebook (2018). URL: https://facebook.com (visited on 09/03/2018).

Fan, Jianqing, Huazhen Lin, Yong Zhou, et al. (2006). "Local partial-likelihood estimation for lifetime data." In: *The Annals of Statistics* 34.1, pp. 290–325.

Fastring, DR and JA Griffith (2009). "Malaria incidence in Nairobi, Kenya and dekadal trends in NDVI and climatic variables." In: *Geocarto International* 24.3, pp. 207–221.

Ferguson, Neil M, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke (2006). "Strategies for mitigating an influenza pandemic." In: *Nature* 442.7101, p. 448.

Fik, Timothy Joseph and Gordon F Mulligan (1998). "Functional form and spatial interaction models." In: *Environment and Planning A* 30.8, pp. 1497–1507.

Findlater, Aidan and Isaac I Bogoch (2018). "Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel." In: *Trends in parasitology*.

Finger, Flavio, Tina Genolet, Lorenzo Mari, Guillaume Constantin de Magny, Noël Magloire Manga, Andrea Rinaldo, and Enrico Bertuzzo (2016). "Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks." In: *Proceedings of the National Academy of Sciences* 113.23, pp. 6421–6426.

Finnie, Thomas JR, Andy South, Ana Bento, Ellie Sherrard-Smith, and Thibaut Jombart (2016). "EpiJSON: A unified data-format for epidemiology." In: *Epidemics* 15, pp. 20–26.

Firth, David (1988). "Multiplicative errors: log-normal or gamma?" In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 266–268.

Fischer, Manfred M (2006). *Spatial analysis and geocomputation: selected essays*. Springer Science & Business Media.

flickr (2018). URL: https://www.flickr.com/ (visited on 09/03/2018).

Forsberg White, Laura and Marcello Pagano (2008). "A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic." In: *Statistics in medicine* 27.16, pp. 2999–3016.

Fraser, Christophe (2007). "Estimating individual and household reproduction numbers in an emerging epidemic." In: *PLoS One* 2.8, e758.

Fratianni, Michele (2007). "The gravity equation in international trade." In:

Funk, Sebastian, Shweta Bansal, Chris T Bauch, Ken TD Eames, W John Edmunds, Alison P Galvani, and Petra Klepac (2015). "Nine challenges in incorporating the dynamics of behaviour in infectious diseases models." In: *Epidemics* 10, pp. 21–25.

Funk, Sebastian, Marcel Salathé, and Vincent AA Jansen (2010). "Modelling the influence of human behaviour on the spread of infectious diseases: a review." In: *Journal of the Royal Society Interface* 7.50, pp. 1247–1256.

Gao, Jianxi, Baruch Barzel, and Albert-László Barabási (2016). "Universal resilience patterns in complex networks." In: *Nature* 530.7590, p. 307.

Garcia, Andres J, Deepa K Pindolia, Kenneth K Lopiano, and Andrew J Tatem (2014). "Modeling internal migration flows in sub-Saharan Africa using census microdata." In: *Migration Studies*, mnu036.

Gardner, Lauren M, Abrar A Chughtai, and C Raina MacIntyre (2016). "Risk of global spread of Middle East respiratory syndrome coronavirus (MERS-CoV) via the air transport network." In: *Journal of travel medicine* 23.6.

Gaudart, Jean, Ousmane Touré, Nadine Dessay, A lassane Dicko, Stéphane Ranque, Loic Forest, Jacques Demongeot, and Ogobara K Doumbo (2009). "Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali." In: *Malaria Journal* 8.1, p. 61.

Gething, Peter W, Peter M Atkinson, Abdisalan M Noor, Priscilla W Gikandi, Simon I Hay, and Mark S Nixon (2007). "A local space–time kriging approach applied to a national outpatient malaria data set." In: *Computers & geosciences* 33.10, pp. 1337–1350.

Gething, Peter W, Abdisalan M Noor, Priscilla W Gikandi, Esther AA Ogara, Simon I Hay, Mark S Nixon, Robert W Snow, and Peter M Atkinson (2006). "Improving imperfect data from health management information systems in Africa using space–time geostatistics." In: *PLoS Med* 3.6, e271.

Geweke, John et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Vol. 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.

Gilks, Walter R, Sylvia Richardson, and David Spiegelhalter (1995). *Markov chain Monte Carlo in practice*. CRC press.

Githeko, Andrew K, Steve W Lindsay, Ulisses E Confalonieri, and Jonathan A Patz (2000). "Climate change and vector-borne diseases: a regional analysis." In: *Bulletin of the World Health Organization* 78.9, pp. 1136–1147.

Gonçalves, Bruno (2015). *Social phenomena : from data analysis to models*. Cham New York: Springer. ISBN: 978-3-319-14010-0.

Gonzalez, Marta C, Cesar A Hidalgo, and Albert-Laszlo Barabasi (2008). "Understanding individual human mobility patterns." In: *Nature* 453.7196, pp. 779–782.

Grange, Louis de, Enrique Fernández, and Joaqun de Cea (2010). "A consolidated model of trip distribution." In: *Transportation Research Part E: Logistics and Transportation Review* 46.1, pp. 61–75.

Grietens, Koen Peeters, Charlotte Gryseels, Susan Dierickx, Melanie Bannister-Tyrrell, Suzan Trienekens, Sambunny Uk, Pisen Phoeuk, Sokha Suon, Srun Set, René Gerrets, et al. (2015a). "Characterizing types of human mobility to inform differential and targeted malaria elimination strategies in Northeast Cambodia." In: *Scientific reports* 5.

Grietens, Koen Peeters, Charlotte Gryseels, Susan Dierickx, Melanie Bannister-Tyrrell, Suzan Trienekens, Sambunny Uk, Pisen Phoeuk, Sokha Suon, Srun Set, René Gerrets, et al. (2015b). "Characterizing types of human mobility to inform differential and targeted malaria elimination strategies in Northeast Cambodia." In: *Scientific reports* 5.

Griffin, Jamie T, T Deirdre Hollingsworth, Lucy C Okell, Thomas S Churcher, Michael White, Wes Hinsley, Teun Bousema, Chris J Drakeley, Neil M Ferguson, Maria-Gloria Basáñez, et al. (2010). "Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies." In: *PLoS Med* 7.8, e1000324.

Halás, Marián, Pavel Klapka, and Petr Kladivo (2014). "Distance-decay functions for daily travel-to-work flows." In: *Journal of Transport Geography* 35, pp. 107–119.

Halloran, M Elizabeth, Alessandro Vespignani, Nita Bharti, Leora R Feldstein, KA Alexander, Matthew Ferrari, Jeffrey Shaman, John M Drake, Travis Porco, Joseph NS Eisenberg, et al. (2014). "Ebola: mobility data." In: *Science* 346.6208, pp. 433–433.

Hamer, Davidson H, Micky Ndhlovu, Dejan Zurovac, Matthew Fox, Kojo Yeboah-Antwi, Pascalina Chanda, Naawa Sipilinyambe, Jonathon L Simon, and Robert W Snow (2007). "Improved diagnostic testing and malaria treatment practices in Zambia." In: *Jama* 297.20, pp. 2227–2231.

Hanel, Rudolf, Bernat Corominas-Murtra, Bo Liu, and Stefan Thurner (2017). "Fitting power-laws in empirical data with estimators that work for all exponents." In: *PloS one* 12.2, e0170920.

Harrington, Laura C, Thomas W Scott, Kriangkrai Lerdthusnee, Russell C Coleman, Adriana Costero, Gary G Clark, James J Jones, Sangvorn Kitthawee, Pattamaporn Kittayapong, Ratana Sithiprasasna, et al. (2005). "Dispersal of the dengue vector Aedes aegypti within and between rural communities." In: *The American journal of tropical medicine and hygiene* 72.2, pp. 209–220.

Hassell, James M, Michael Begon, Melissa J Ward, and Eric M Fèvre (2017). "Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface." In: *Trends in ecology & evolution* 32.1, pp. 55–67.

Hawelka, Bartosz, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti (2014). "Geo-located Twitter as proxy for global mobility patterns." In: *Cartography and Geographic Information Science* 41.3, pp. 260–271.

Hay, Simon I, Dylan B George, Catherine L Moyes, and John S Brownstein (2013). "Big data opportunities for global infectious disease surveillance." In: *PLoS medicine* 10.4, e1001413.

Hay, Simon I and Robert W Snow (2006). "The Malaria Atlas Project: developing global maps of malaria risk." In: *PLoS medicine* 3.12, e473.

Haynes, Kingsley E, A Stewart Fotheringham, et al. (1984). *Gravity and spatial interaction models*. Vol. 2. Sage Beverly Hills, CA.

Heffernan, JM, RJ Smith, and LM Wahl (2005). "Perspectives on the basic reproductive ratio." In: *Journal of the Royal Society Interface* 2.4, pp. 281–293.

Henry, Sabine, Paul Boyle, and Eric F Lambin (2003). "Modelling inter-provincial migration in Burkina Faso, West Africa: the role of socio-demographic and environmental factors." In: *Applied Geography* 23.2, pp. 115–136.

Hering, Amanda S, Cynthia L Bell, and Marc G Genton (2009). "Modeling spatio-temporal wildfire ignition point patterns." In: *Environmental and Ecological Statistics* 16.2, pp. 225–250.

Heymann, David L and Guenael Rodier (2004). "SARS: lessons from a new disease." In: *Learning from SARS: Preparing for the next disease outbreak*, pp. 234–245.

Hoek, W Van Der, F Konradsen, D Perera, Priyanie H Amerasinghe, and Felix P Amerasinghe (1997). "Correlation between rainfall and malaria in the dry zone of Sri Lanka." In: *Annals of Tropical Medicine & Parasitology* 91.8, pp. 945–949.

Hoffman, Matthew D and Andrew Gelman (2014). "The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." In: *Journal of Machine Learning Research* 15.1, pp. 1593–1623.

Höhle, Michael (2009). "Additive-Multiplicative Regression Models for Spatio-Temporal Epidemics." In: *Biometrical journal* 51.6, pp. 961–978.

Holmes, Alison H, Luke SP Moore, Arnfinn Sundsfjord, Martin Steinbakk, Sadie Regmi, Abhilasha Karkey, Philippe J Guerin, and Laura JV Piddock (2016). "Understanding the mechanisms and drivers of antimicrobial resistance." In: *The Lancet* 387.10014, pp. 176–187.

House, Thomas, Marc Baguelin, Albert Jan Van Hoek, Peter J White, Zia Sadique, Ken Eames, Jonathan M Read, Niel Hens, Alessia Melegaro, W John Edmunds, et al. (2011). "Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic." In: *Proceedings of the Royal Society of London B: Biological Sciences* 278.1719, pp. 2753–2760.

Hu, Hao, Karima Nigmatulina, and Philip Eckhoff (2013). "The scaling of contact rates with population density for the infectious disease models." In: *Mathematical biosciences* 244.2, pp. 125–134.

Huang, Qunying and David WS Wong (2015). "Modeling and visualizing regular human mobility patterns with uncertainty: An example using Twitter data." In: *Annals of the Association of American Geographers* 105.6, pp. 1179–1197.

Hufnagel, Lars, Dirk Brockmann, and Theo Geisel (2004). "Forecast and control of epidemics in a globalized world." In: *Proceedings of the National Academy of Sciences of the United States of America* 101.42, pp. 15124–15129.

Hui, David S, Stanley Perlman, and Alimuddin Zumla (2015). "Spread of MERS to South Korea and China." In: *The lancet Respiratory medicine* 3.7, pp. 509–510.

Hung, Lee Shiu (2003). "The SARS epidemic in Hong Kong: what lessons have we learned?" In: *Journal of the Royal Society of Medicine* 96.8, pp. 374–378.

Intelligence, GSMA (2016). *The Mobile Economy 2016*.

Jancloes, Michel, Madeleine Thomson, Maria Máñez Costa, Chris Hewitt, Carlos Corvalan, Tufa Dinku, Rachel Lowe, and Mary Hayden (2014). "Climate services to improve public health." In: *International journal of environmental research and public health* 11.5, pp. 4555–4559.

Jandarov, Roman, Murali Haran, Ottar Bjørnstad, and Bryan Grenfell (2014). "Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 63.3, pp. 423–444.

Jepson, WF, A Moutia, and C Courtois (1947). "The malaria problem in Mauritius: the bionomics of Mauritian anophelines." In: *Bulletin of entomological research* 38.01, pp. 177–208.

Jiang, Chao Qiang, Justin Lessler, Lina Kim, Kin On Kwok, Jonathan M Read, Shuying Wang, Lijiu Tan, Marisa Hast, Huachen Zhu, Yi Guan, et al. (2016). "Cohort Profile: A study of influenza immunity in the urban and rural Guangzhou region of China: the Fluscape Study." In: *International journal of epidemiology* 46.2, e16–e16.

Jiang, Shan, Joseph Ferreira, and Marta C González (2017). "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore." In: *IEEE Transactions on Big Data* 3.2, pp. 208–219.

Kang, Chaogui, Yu Liu, Diansheng Guo, and Kun Qin (2015). "A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint." In: *PloS one* 10.11, e0143500.

Karamshuk, Dmytro, Chiara Boldrini, Marco Conti, and Andrea Passarella (2011). "Human mobility models for opportunistic networks." In: *IEEE Communications Magazine* 49.12, pp. 157–165.

Kau, James B and CF Sirmans (1979). "The functional form of the gravity model." In: *International Regional Science Review* 4.2, pp. 127–136.

Kaufmann, Christian and Hans Briegel (2004). "Flight performance of the malaria vectors Anopheles gambiae and Anopheles atroparvus." In: *Journal of vector ecology* 29, pp. 140–153.

Kazembe, Lawrence N, Immo Kleinschmidt, Timothy H Holtz, and Brian L Sharp (2006). "Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data." In: *International Journal of Health Geographics* 5.1, p. 41.

Keeling, Matt J and Ken TD Eames (2005). "Networks and epidemic models." In: *Journal of the Royal Society Interface* 2.4, pp. 295–307.

Keiser, Jennifer, Jürg Utzinger, Marcia Caldas De Castro, Thomas A Smith, Marcel Tanner, and Burton H Singer (2004). "Urbanization in sub-saharan Africa and implication for malaria control." In: *The American journal of tropical medicine and hygiene* 71.2 suppl, pp. 118–127.

Kelsall, Julia and Jonathan Wakefield (2002). "Modeling spatial variation in disease risk: a geostatistical approach." In: *Journal of the American Statistical Association* 97.459, pp. 692–701.

Kenah, Eben, Marc Lipsitch, and James M Robins (2008). "Generation interval contraction and epidemic data analysis." In: *Mathematical biosciences* 213.1, pp. 71–79.

Klassen, Sarah, Jonathan Weed, and Damian Evans (2018). "Semi-supervised machine learning approaches for predicting the chronology of archaeological sites: A case study of temples from medieval Angkor, Cambodia." In: *PloS one* 13.11, e0205649.

Knorr-Held, Leonhard (1999). "Bayesian modelling of inseparable space-time variation in disease risk." In:

Kotsiantis, Sotiris B, I Zaharakis, and P Pintelas (2007). "Supervised machine learning: A review of classification techniques." In: *Emerging artificial intelligence applications in computer engineering* 160, pp. 3–24.

Kottas, Athanasios, Jason A Duan, and Alan E Gelfand (2008). "Modeling disease incidence data with spatial and spatio temporal Dirichlet process mixtures." In: *Biometrical Journal* 50.1, pp. 29–42.

Kramer, Andrew M, J Tomlin Pulliam, Laura W Alexander, Andrew W Park, Pejman Rohani, and John M Drake (2016). "Spatial spread of the West Africa Ebola epidemic." In: *Open Science* 3.8, p. 160294.

Krenn, Patricia J, Sylvia Titze, Pekka Oja, Andrew Jones, and David Ogilvie (2011). "Use of global positioning systems to study physical activity and the environment: a systematic review." In: *American journal of preventive medicine* 41.5, pp. 508–515.

Lafferty, Kevin D (2009). "The ecology of climate change and infectious diseases." In: *Ecology* 90.4, pp. 888–900.

LandDAAC MODIS (2018). URL: http://iridl.ldeo.columbia.edu/SOURCES/.USGS/.LandDAAC/.MODIS/.version_005/.SAF/ (visited on 09/03/2018).

Lau, Joseph TF, Xilin Yang, Ellie Pang, HY Tsui, Eric Wong, and Yun Kwok Wing (2005). "SARS-related perceptions in Hong Kong." In: *Emerging infectious diseases* 11.3, p. 417.

Lau, Max SY, Benjamin Douglas Dalziel, Sebastian Funk, Amanda McClelland, Amanda Tiffany, Steven Riley, C Jessica E Metcalf, and Bryan T Grenfell (2017). "Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic." In: *Proceedings of the National Academy of Sciences* 114.9, pp. 2337–2342.

Lawson, Andrew B (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.

Lawson, Andrew B and Petra Leimich (2000a). "Approaches to the space-time modelling of infectious disease behaviour." In: *Mathematical Medicine and Biology* 17.1, pp. 1–13.

Lawson, Andrew B and Petra Leimich (2000b). "Approaches to the space-time modelling of infectious disease behaviour." In: *Mathematical Medicine and Biology: A Journal of the IMA* 17.1, pp. 1–13.

Lee, Duncan (2011). "A comparison of conditional autoregressive models used in Bayesian disease mapping." In: *Spatial and Spatio-temporal Epidemiology* 2.2, pp. 79–89.

Lee, Duncan and Andrew Lawson (2014). "Cluster detection and risk estimation for spatio-temporal health data." In: *arXiv preprint arXiv:1408.1191*.

Lee, Duncan, Alastair Rushworth, and Gary Napier (2015). "CARBayesST: An R Package for Spatio-temporal Areal Unit Modelling with Conditional Autoregressive Priors." In:

Lee, Elizabeth C, Jason M Asher, Sandra Goldlust, John D Kraemer, Andrew B Lawson, and Shweta Bansal (2016). "Mind the scales: Harnessing spatial big data for infectious disease surveillance and inference." In: *The Journal of infectious diseases* 214.suppl_4, S409–S413.

Lee, Jonggul, Bo Youl Choi, and Eunok Jung (2018). "Metapopulation model using commuting flow for national spread of the 2009 H1N1 influenza virus in the Republic of Korea." In: *Journal of Theoretical Biology* 454, pp. 320–329.

Leone, Statistics Sierra (2015). *2015 Population and Housing Census*. https://www.statistics.sl/wp-content/uploads/2017/01/final-results_-2015_population_and_housing_census.pdf.

Leroux, Brian G, Xingye Lei, and Norman Breslow (2000). "Estimation of disease rates in small areas: a new mixed model for spatial dependence." In: *Statistical models in epidemiology, the environment, and clinical trials*. Springer, pp. 179–191.

LeSage, James P and Carlos Llano (2016). "A spatial interaction model with spatially structured origin and destination effects." In: *Spatial Econometric Interaction Modelling*. Springer, pp. 171–197.

LeSage, James P and R Kelley Pace (2008). "Spatial econometric modeling of origin-destination flows." In: *Journal of Regional Science* 48.5, pp. 941–967.

Lessler, Justin, Andrew S Azman, M Kate Grabowski, Henrik Salje, and Isabel Rodriguez-Barraquer (2016). "Trends in the mechanistic and dynamic modeling of infectious diseases." In: *Current Epidemiology Reports* 3.3, pp. 212–222.

Lessler, Justin, Derek AT Cummings, Jonathan M Read, Shuying Wang, Huachen Zhu, Gavin JD Smith, Yi Guan, Chao Qiang Jiang, and Steven Riley (2011). "Location-specific patterns of exposure to recent pre-pandemic strains of influenza A in southern China." In: *Nature communications* 2, p. 423.

Li, Xinhai, Huidong Tian, Dejian Lai, and Zhibin Zhang (2011). "Validation of the gravity model in predicting the global spread of influenza." In: *International journal of environmental research and public health* 8.8, pp. 3134–3143.

Li, Ye, Patrick Brown, Dionne C Gesink, and Håvard Rue (2012). "Log Gaussian Cox processes and spatially aggregated disease incidence data." In: *Statistical methods in medical research* 21.5, pp. 479–507.

Liang, Xiao, Jichang Zhao, Li Dong, and Ke Xu (2013). "Unraveling the origin of exponential law in intra-urban human mobility." In: *Scientific reports* 3.

Lin, DY and Zhiliang Ying (1995). "Semiparametric analysis of general additive-multiplicative hazard models for counting processes." In: *The annals of Statistics*, pp. 1712–1734.

Lin, Feng-Chang and Jun Zhu (2012). "Additive hazards regression and partial likelihood estimation for ecological monitoring data across space." In: *Statistics and its Interface* 5.2.

Lindblade, Kim A, Edward D Walker, Ambrose W Onapa, Justus Katungu, and Mark L Wilson (2000). "Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda." In: *Tropical Medicine & International Health* 5.4, pp. 263–274.

Lindsay, Steven W, René Bødker, Robert Malima, Hamisi A Msangeni, and Wiliam Kisinza (2000). "Effect of 1997–98 EI Niño on highland malaria in Tanzania." In: *The Lancet* 355.9208, pp. 989–990.

LINK (2018). URL: https://www.linkmalaria.org/ (visited on 09/03/2018).

Loog, Liisa, Marta Mirazón Lahr, Mirna Kovacevic, Andrea Manica, Anders Eriksson, and Mark G Thomas (2017). "Estimating mobility using sparse data: Application to human genetic variation." In: *Proceedings of the National Academy of Sciences* 114.46, pp. 12213–12218.

Lowe, Rachel, Trevor C Bailey, David B Stephenson, Richard J Graham, Caio AS Coelho, Marilia Sa Carvalho, and Christovam Barcellos (2011). "Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil." In: *Computers & Geosciences* 37.3, pp. 371–381.

Lowe, Rachel, Trevor C Bailey, David B Stephenson, Tim E Jupp, Richard J Graham, Christovam Barcellos, and Marilia Sá Carvalho (2013a). "The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil." In: *Statistics in medicine* 32.5, pp. 864–883.

Lowe, Rachel, Christovam Barcellos, Caio AS Coelho, Trevor C Bailey, Giovanini Evelim Coelho, Richard Graham, Tim Jupp, Walter Massa Ramalho, Marilia Sá Carvalho, David B Stephenson, et al. (2014a). "Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts." In: *The Lancet infectious diseases* 14.7, pp. 619–626.

Lowe, Rachel, Christovam Barcellos, Caio AS Coelho, Trevor C Bailey, Giovanini Evelim Coelho, Richard Graham, Tim Jupp, Walter Massa Ramalho, Marilia Sá Carvalho, David B Stephenson, et al. (2014b). "Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts." In: *The Lancet Infectious Diseases* 14.7, pp. 619–626.

Lowe, Rachel, James Chirombo, and Adrian M Tompkins (2013b). "Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi." In: *Malaria journal* 12.1, p. 1.

Lowe, Rachel, Caio AS Coelho, Christovam Barcellos, Marilia Sa Carvalho, Rafael De Castro Catao, Giovanini E Coelho, Walter Massa Ramalho, Trevor C Bailey, David B Stephenson, and Xavier Rodo (2016). "Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil." In: *Elife* 5, e11285.

Lowe, Rachel, Anna M Stewart-Ibarra, Desislava Petrova, Markel Garcıa-Dıez, Mercy J Borbor-Cordova, Raúl Mejıa, Mary Regato, and Xavier Rodó (2017). "Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador." In: *The Lancet Planetary Health* 1.4, e142–e151.

Lyon, Bradfield, Tufa Dinku, Anita Raman, and Madeleine C Thomson (2017). "Temperature suitability for malaria climbing the Ethiopian Highlands." In: *Environmental Research Letters* 12.6.

Ma, Haijun, Bradley P Carlin, et al. (2007). "Bayesian multivariate areal wombling for multiple disease boundary analysis." In: *Bayesian analysis* 2.2, pp. 281–302.

Mabey, David, Rosanna W Peeling, Andrew Ustianowski, and Mark D Perkins (2004). "Tropical infectious diseases: diagnostics for the developing world." In: *Nature Reviews Microbiology* 2.3, pp. 231–240.

MacLachlan, Ian (2011). "Quantitative Models for Geographical Analysis Additional Required Reading for Geography 3235 (Fourth Revision)." In:

Macro, ORC and National Statistical Office (2005). "Malawi demographic and health survey 2004." In: *Zomba, Malawi: National Statistical Office*.

Malawi Health Management Information System (HMIS) (n.d.). http://live.hispmalawi.org.mw/.

Malawi, MDG (2015). "Malawi MDG Endline Survey 2014." In: *Zomba: National Statistical Office*.

Manning, Willard G and John Mullahy (2001). "Estimating log models: to transform or not to transform?" In: *Journal of health economics* 20.4, pp. 461–494.

Mao, Huina, Xin Shuai, Yong-Yeol Ahn, and Johan Bollen (2015). "Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire." In: *EPJ Data Science* 4.1, pp. 1–16.

Marshall, John M, Adam Bennett, Samson S Kiware, and Hugh JW Sturrock (2016). "The hitchhiking parasite: why human movement matters to malaria transmission and what we can do about it." In: *Trends in parasitology* 32.10, pp. 752–755.

Marshall, John M, Sean L Wu, Samson S Kiware, Micky Ndhlovu, André Lin Ouédraogo, Mahamoudou B Touré, Hugh J Sturrock, Azra C Ghani, Neil M Ferguson, et al. (2018). "Mathematical models of human mobility of relevance to malaria transmission in Africa." In: *Scientific reports* 8.1, p. 7713.

Martens, Pim and Lisbeth Hall (2000). "Malaria on the move: human population movement and malaria transmission." In: *Emerging infectious diseases* 6.2, p. 103.

Martinussen, Torben and Thomas H Scheike (2002). "A flexible additive multiplicative hazard model." In: *Biometrika* 89.2, pp. 283–298.

Martınez-Zarzoso, Inmaculada (2013). "The log of gravity revisited." In: *Applied Economics* 45.3, pp. 311–327.

Masucci, A Paolo, Joan Serras, Anders Johansson, and Michael Batty (2013). "Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows." In: *Physical Review E* 88.2, p. 022812.

McCloskey, Brian, Osman Dar, Alimuddin Zumla, and David L Heymann (2014). "Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread." In: *The Lancet infectious diseases* 14.10, pp. 1001–1010.

McCullagh, Peter and John A Nelder (1989). *Generalized linear models*. Vol. 37. CRC press.

Meekan, Mark G, Carlos M Duarte, Juan Fernández-Gracia, Michele Thums, Ana MM Sequeira, Rob Harcourt, and Vıctor M Eguluz (2017). "The ecology of human mobility." In: *Trends in ecology & evolution* 32.3, pp. 198–210.

Meyer, Sebastian, Johannes Elias, and Michael Höhle (2012). "A space–time conditional intensity model for invasive meningococcal disease occurrence." In: *Biometrics* 68.2, pp. 607–616.

Meyer, Sebastian and Leonhard Held (2014). "Power-law models for infectious disease spread." In: *The Annals of Applied Statistics* 8.3, pp. 1612–1639.

Meyer, Sebastian and Leonhard Held (2017). "Incorporating social contact data in spatio-temporal models for infectious disease spread." In: *Biostatistics* 18.2, pp. 338–351.

Meyer, Sebastian, Leonhard Held, et al. (2014a). "Power-law models for infectious disease spread." In: *The Annals of Applied Statistics* 8.3, pp. 1612–1639.

Meyer, Sebastian, Leonhard Held, et al. (2014b). "Power-law models for infectious disease spread." In: *The Annals of Applied Statistics* 8.3, pp. 1612–1639.

Meyer, Sebastian, Leonhard Held, and Michael Höhle (2017). "Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package surveillance." In: *Journal of Statistical Software, Articles* 77.11, pp. 1–55.

Ministry of Health (MOH) (2011). "Guidelines for use of malaria rapid diagnostic tests (mRDTs) in Malawi." In:

Ministry of Health (MOH) (2015). *Malawi Health Sector Strategic Plan 2011-2016*. Lilongwe, Malawi.

Møller, Jesper, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen (1998). "Log gaussian cox processes." In: *Scandinavian journal of statistics* 25.3, pp. 451–482.

Molodecky, Natalie A, Isobel M Blake, Kathleen M O'Reilly, Mufti Zubair Wadood, Rana M Safdar, Amy Wesolowski, Caroline O Buckee, Ananda S Bandyopadhyay, Hiromasa Okayasu, and Nicholas C Grassly (2017). "Risk factors and short-term projections for serotype-1 poliomyelitis incidence in Pakistan: A spatiotemporal analysis." In: *PLoS medicine* 14.6, e1002323.

Morens, David M and Anthony S Fauci (2013). "Emerging infectious diseases: threats to human health and global stability." In: *PLoS pathogens* 9.7, e1003467.

Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. (2008). "Social contacts and mixing patterns relevant to the spread of infectious diseases." In: *PLoS medicine* 5.3, e74.

Musmeci, Fabio and David Vere-Jones (1992). "A space-time clustering model for historical earthquakes." In: *Annals of the Institute of Statistical Mathematics* 44.1, pp. 1–11.

Nabarro, D (1999). "Roll Back Malaria." In: *Parassitologia* 41.1-3, pp. 501–504.

Nabarro, David N and Elizabeth M Tayler (1998). "The" roll back malaria" campaign." In: *Science* 280.5372, pp. 2067–2068.

National Statistical Office (NSO) (2008). *Malawi Population Projections*. Zomba: NSO.

National Statistical Office (NSO) (2010a). *Malawi Population Projections 2008*. Zomba.

National Statistical Office (NSO) (2010b). *Welfare Monitoring Survey 2009*. Zomba: NSO.

National Statistical Office (NSO) (2012). *Integrated Household Survey 2010-2011*. Zomba,Malawi.

National Statistical Office (NSO) and ICF Macro (2011). *Malawi Demographic and Health Survey 2010*. Zomba, Malawi & Calverton, Maryland, USA.

Neiderud, Carl-Johan (2015). "How urbanization affects the epidemiology of emerging infectious diseases." In: *Infection ecology & epidemiology* 5.1, p. 27060.

Nelder, John A and Roger Mead (1965). "A simplex method for function minimization." In: *The computer journal* 7.4, pp. 308–313.

Ngongondo, Cosmo, Chong-Yu Xu, Lars Gottschalk, and Berhanu Alemaw (2011). "Evaluation of spatial and temporal characteristics of rainfall in Malawi: a case of data scarce region." In: *Theoretical and applied climatology* 106.1-2, pp. 79–93.

Nicholson, SE, D Klotter, and G Chavula (2014). "A detailed rainfall climatology for Malawi, Southern Africa." In: *International Journal of Climatology* 34.2, pp. 315–325.

Nishiura, H and G Chowell (2014). "Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014." In: *Euro Surveill* 19.36, p. 20894.

Nkegbe, Paul Kwame, Naasegnibe Kuunibe, and Samuel Sekyi (2017). "Poverty and malaria morbidity in the Jirapa District of Ghana: A count regression approach." In: *Cogent Economics & Finance* 5.1, p. 1293472.

Nkumama, Irene N, Wendy P O'Meara, and Faith HA Osier (2017). "Changes in malaria epidemiology in Africa and new challenges for elimination." In: *Trends in parasitology* 33.2, pp. 128–140.

NMCP, National Malaria Control Programme (2011). *Malawi Malaria Indicator Survey 2010*. Lilongwe,Malawi: NMCP.

NMCP, National Malaria Control Programme and ICF (2018). "Malawi Malaria Indicator Survey 2017." In:

NOAA (2018). URL: https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.GHCN_CAMS/.gridded/.deg0p5 (visited on 09/03/2018).

NSO (2012). "Intergrated Household Survey (IHS 3) Report." In: p. 12.

Ogata, Yosihiko (1998). "Space-time point-process models for earthquake occurrences." In: *Annals of the Institute of Statistical Mathematics* 50.2, pp. 379–402.

OrganizationWHO, World Health (2015). *Global technical strategy for malaria 2016-2030*. World Health Organization.

Paaijmans, Krijn P, Andrew F Read, and Matthew B Thomas (2009). "Understanding the link between malaria risk and climate." In: *Proceedings of the National Academy of Sciences* 106.33, pp. 13844–13849.

Paaijmans, Krijn P, Moses O Wandago, Andrew K Githeko, and Willem Takken (2007). "Unexpected high losses of Anopheles gambiae larvae due to rainfall." In: *PLoS One* 2.11, e1146.

Palchykov, Vasyl, Marija Mitrović, Hang-Hyun Jo, Jari Saramäki, and Raj Kumar Pan (2014). "Inferring human mobility using communication patterns." In: *Scientific reports* 4.

Panigutti, Cecilia, Michele Tizzoni, Paolo Bajardi, Zbigniew Smoreda, and Vittoria Colizza (2017). "Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models." In: *Royal Society open science* 4.5, p. 160950.

Pappalardo, Luca, Dino Pedreschi, Zbigniew Smoreda, and Fosca Giannotti (2015). "Using big data to study the link between human mobility and socio-economic development." In: *Big Data (Big Data), 2015 IEEE International Conference on*. IEEE, pp. 871–878.

Parashar, Umesh D and Larry J Anderson (2004). *Severe acute respiratory syndrome: review and lessons of the 2003 outbreak*.

Parham, Paul Edward and Edwin Michael (2010). "Modeling the effects of weather and climate change on malaria transmission." In: *Environmental health perspectives* 118.5, p. 620.

Pati, Debdeep and David B Dunson (2014). "Bayesian nonparametric regression with varying residual density." In: *Annals of the Institute of Statistical Mathematics* 66.1, pp. 1–31.

Paulo, Rui (2005). "Default priors for Gaussian processes." In: *Annals of Statistics*, pp. 556–582.

Paz-Soldan, Valerie A, Robert C Reiner Jr, Amy C Morrison, Steven T Stoddard, Uriel Kitron, Thomas W Scott, John P Elder, Eric S Halsey, Tadeusz J Kochel, Helvio Astete, et al. (2014). "Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru." In: *PLoS neglected tropical diseases* 8.6, e2888.

Pellegrini, Pasquale A and A Stewart Fotheringham (2002). "Modelling spatial choice: a review and synthesis in a migration context." In: *Progress in human geography* 26.4, pp. 487–510.

Perez-Saez, Javier, Lorenzo Mari, Enrico Bertuzzo, Renato Casagrandi, Susanne H Sokolow, Giulio A De Leo, Theophile Mande, Natalie Ceperley, Jean-Marc Froehlich, Mariam Sou, et al. (2015). "A theoretical analysis of the geography of schistosomiasis in Burkina Faso highlights the roles of human mobility and water resources development in disease transmission." In: *PLoS neglected tropical diseases* 9.10, e0004127.

Porojan, Anca (2001). "Trade flows and spatial effects: the gravity model revisited." In: *Open economies review* 12.3, pp. 265–280.

Prem, Kiesha, Alex R Cook, and Mark Jit (2017). "Projecting social contact matrices in 152 countries using contact surveys and demographic data." In: *PLoS computational biology* 13.9, e1005697.

Prothero, Ralph Mansell (2001). "Migration and malaria risk." In: *Health, Risk & Society* 3.1, pp. 19–38.

Ravenstein, Ernst Georg (1885). "The laws of migration." In: *Journal of the statistical society of London* 48.2, pp. 167–235.

Read, JM, WJ Edmunds, S Riley, J Lessler, and DAT Cummings (2012). "Close encounters of the infectious kind: methods to measure social mixing behaviour." In: *Epidemiology & Infection* 140.12, pp. 2117–2130.

Read, Jonathan M, Peter J Diggle, James Chirombo, Tom Solomon, and Matthew Baylis (2015). "Effectiveness of screening for Ebola at airports." In: *The Lancet* 385.9962, pp. 23–24.

Read, Jonathan M, Justin Lessler, Steven Riley, Shuying Wang, Li Jiu Tan, Kin On Kwok, Yi Guan, Chao Qiang Jiang, and Derek AT Cummings (2014). "Social mixing patterns in rural and urban areas of southern China." In: *Proceedings of the Royal Society of London B: Biological Sciences* 281.1785, p. 20140268.

Reiner, Robert C, Steven T Stoddard, and Thomas W Scott (2014). "Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal." In: *Epidemics* 6, pp. 30–36.

Reyburn, Hugh, Redepmta Mbatia, Chris Drakeley, Ilona Carneiro, Emmanuel Mwakasungula, Ombeni Mwerinde, Kapalala Saganda, John Shao, Andrew Kitua, Raimos Olomi, et al. (2004). "Overdiagnosis of malaria in patients with severe febrile illness in Tanzania: a prospective study." In: *Bmj* 329.7476, p. 1212.

Riley, Steven (2007). "Large-scale spatial-transmission models of infectious disease." In: *Science* 316.5829, pp. 1298–1301.

Riley, Steven and Neil M Ferguson (2006). "Smallpox transmission and control: spatial dynamics in Great Britain." In: *Proceedings of the National Academy of Sciences* 103.33, pp. 12637–12642.

Robert, Vincent, Kate Macintyre, Joseph Keating, Jean-Francois Trape, Jean-Bernard Duchemin, McWilson Warren, and John C Beier (2003). "Malaria transmission in urban sub-Saharan Africa." In: *The American journal of tropical medicine and hygiene* 68.2, pp. 169–176.

Roberts, Mick, Viggo Andreasen, Alun Lloyd, and Lorenzo Pellis (2015). "Nine challenges for deterministic epidemic models." In: *Epidemics* 10, pp. 49–53.

Ruan, Zhongyuan, Chaoqing Wang, Pak Ming Hui, and Zonghua Liu (2015). "Integrated travel network model for studying epidemics: Interplay between journeys and epidemic." In: *Scientific reports* 5.

Rue, Havard and Leonhard Held (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

Rushworth, Alastair, Duncan Lee, and Richard Mitchell (2014). "A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London." In: *Spatial and spatio-temporal epidemiology* 10, pp. 29–38.

Rushworth, Alastair, Duncan Lee, and Christophe Sarran (2016). "An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

Sarzynska, Marta, Oyita Udiani, and Na Zhang (2013). "A study of gravity-linked metapopulation models for the spatial spread of dengue fever." In: *arXiv preprint arXiv:1308.4589*.

Sattenspiel, Lisa and Alun Lloyd (2009). *The geographic spread of infectious diseases: models and applications*. Princeton University Press.

Schicker, Rebekah Stewart, Neway Hiruy, Berhanu Melak, Woyneshet Gelaye, Belay Bezabih, Rob Stephenson, Amy E Patterson, Zerihun Tadesse, Paul M Emerson, Frank O Richards Jr, et al. (2015). "A venue-based survey of malaria, anemia and mobility patterns among migrant farm workers in Amhara region, Ethiopia." In: *PloS one* 10.11, e0143829.

Searle, Kelly M, Jailos Lubinda, Harry Hamapumbu, Timothy M Shields, Frank C Curriero, David L Smith, Philip E Thuma, and William J Moss (2017). "Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia." In: *Royal Society open science* 4.5, p. 170046.

Sen, Ashish and Tony Smith (2012). *Gravity models of spatial interaction behavior*. Springer Science & Business Media.

Shultz, James M, Zelde Espinel, Maria Espinola, and Andreas Rechkemmer (2016). "Distinguishing epidemiological features of the 2013–2016 West Africa Ebola virus disease outbreak." In: *Disaster Health* 3.3, pp. 78–88.

Shuman, Emily K (2010). "Global climate change and infectious diseases." In: *New England Journal of Medicine* 362.12, pp. 1061–1063.

Silal, Sheetal Prakash, Francesca Little, Karen Irma Barnes, and Lisa Jane White (2015). "Hitting a moving target: a model for malaria elimination in the presence of population movement." In: *PloS one* 10.12, e0144990.

Siliverstovs, Boriss and Dieter Schumacher (2009). "Estimating gravity equations: to log or not to log?" In: *Empirical Economics* 36.3, pp. 645–669.

Silva, JMC Santos and Silvana Tenreyro (2006). "The log of gravity." In: *The Review of Economics and statistics* 88.4, pp. 641–658.

Silva, JMC Santos and Silvana Tenreyro (2011). "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator." In: *Economics Letters* 112.2, pp. 220–222.

Simini, Filippo, Marta C González, Amos Maritan, and Albert-László Barabási (2012). "A universal model for mobility and migration patterns." In: *Nature* 484.7392, pp. 96–100.

Snow, Robert W, Marlies Craig, Uwe Deichmann, and Karen Marsh (1999). "Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population." In: *Bulletin of the World Health Organization* 77.8, p. 624.

Snow, Robert W, Benn Sartorius, David Kyalo, Joseph Maina, Punam Amratia, Clara W Mundia, Philip Bejon, and Abdisalan M Noor (2017). "The prevalence of Plasmodium falciparum in sub-Saharan Africa since 1900." In: *Nature* 550.7677, p. 515.

Song, Chaoming, Tal Koren, Pu Wang, and Albert-László Barabási (2010). "Modelling the scaling properties of human mobility." In: *Nature Physics* 6.10, pp. 818–823.

Song, Xuan, Quanshi Zhang, Yoshihide Sekimoto, Ryosuke Shibasaki, Nicholas Jing Yuan, and Xing Xie (2017). "Prediction and simulation of human mobility following natural disasters." In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 8.2, p. 29.

Sorichetta, Alessandro, Tom J Bird, Nick W Ruktanonchai, Elisabeth zu Erbach-Schoenberg, Carla Pezzulo, Natalia Tejedor, Ian C Waldock, Jason D Sadler, Andres J Garcia, Luigi Sedda, et al. (2016). "Mapping internal connectivity through human migration in malaria endemic countries." In: *Scientific data* 3, p. 160066.

Sorichetta, Alessandro, Graeme M Hornby, Forrest R Stevens, Andrea E Gaughan, Catherine Linard, and Andrew J Tatem (2015). "High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020." In: *Scientific data* 2, p. 150045.

Soto, SM (2009). "Human migration and infectious diseases." In: *Clinical Microbiology and Infection* 15.s1, pp. 26–28.

Statistics, Liberia Institute of and Geo-Information Services (LISGIS (n.d.).

Statistique, Institut National de la (2017). *Population and Housing Census*. URL: http://www.stat-guinee.org/.

Stein, Richard A (2011). "Super-spreaders in infectious diseases." In: *International Journal of Infectious Diseases* 15.8, e510–e513.

Stewart-Ibarra, Anna M and Rachel Lowe (2013). "Climate and non-climate drivers of dengue epidemics in southern coastal Ecuador." In: *The American journal of tropical medicine and hygiene* 88.5, pp. 971–981.

Stoddard, Steven T, Brett M Forshey, Amy C Morrison, Valerie A Paz-Soldan, Gonzalo M Vazquez-Prokopec, Helvio Astete, Robert C Reiner, Stalin Vilcarromero, John P Elder, Eric S Halsey, et al. (2013). "House-to-house human movement drives dengue virus transmission." In: *Proceedings of the National Academy of Sciences* 110.3, pp. 994–999.

Stoddard, Steven T, Amy C Morrison, Gonzalo M Vazquez-Prokopec, Valerie Paz Soldan, Tadeusz J Kochel, Uriel Kitron, John P Elder, and Thomas W Scott (2009). "The role of human movement in the transmission of vector-borne pathogens." In: *PLoS neglected tropical diseases* 3.7, e481.

Stouffer, Samuel A (1940). "Intervening opportunities: a theory relating mobility and distance." In: *American sociological review* 5.6, pp. 845–867.

Stresman, Gillian H (2010). "Beyond temperature and precipitation: ecological risk factors that modify malaria transmission." In: *Acta Tropica* 116.3, pp. 167–172.

Svoboda, Tomislav, Bonnie Henry, Leslie Shulman, Erin Kennedy, Elizabeth Rea, Wil Ng, Tamara Wallington, Barbara Yaffe, Effie Gournis, Elisa Vicencio, et al. (2004). "Public health measures to control the spread of the severe acute respiratory syndrome during the outbreak in Toronto." In: *New England Journal of Medicine* 350.23, pp. 2352–2361.

Tatem, Andrew J, Zhuojie Huang, Clothilde Narib, Udayan Kumar, Deepika Kandula, Deepa K Pindolia, David L Smith, Justin M Cohen, Bonita Graupe, Petrina Uusiku, et al. (2014). "Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning." In: *Malaria journal* 13.1, p. 1.

Tatem, Andrew J, David J Rogers, and SI Hay (2006). "Global transport networks and infectious disease spread." In: *Advances in parasitology* 62, pp. 293–343.

Team, Stan Developers (2015). "Stan Modeling Language: User's Guide and Reference Manual." In: *Version 2.12*.

Teklehaimanot, Hailay D, Marc Lipsitch, Awash Teklehaimanot, and Joel Schwartz (2004). "Weather-based prediction of Plasmodium falciparum malaria in epidemic-prone regions of Ethiopia I. Patterns of lagged weather effects reflect biological mechanisms." In: *Malaria journal* 3.1, p. 41.

Thomson, Madeleine (2018). *IRI*. URL: http://www.climate-services.org/wp-content/uploads/2017/03/S14_01_Thomson_etal.pdf (visited on 09/24/2018).

Thomson, Madeleine C, Simon J Mason, Thandie Phindela, and Stephen J Connor (2005). "Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana." In: *The American journal of tropical medicine and hygiene* 73.1, pp. 214–221.

Thomson, MC, FJ Doblas-Reyes, SJ Mason, R Hagedorn, SJ Connor, T Phindela, AP Morse, and TN Palmer (2006). "Malaria early warnings based on seasonal climate forecasts from multi-model ensembles." In: *Nature* 439.7076, pp. 576–579.

Tian, Huaiyu, Zhe Sun, Nuno Rodrigues Faria, Jing Yang, Bernard Cazelles, Shanqian Huang, Bo Xu, Qiqi Yang, Oliver G Pybus, and Bing Xu (2017). "Increasing airline travel may facilitate co-circulation of multiple dengue virus serotypes in Asia." In: *PLoS neglected tropical diseases* 11.8, e0005694.

Tilahun, Nebiyou and David Levinson (2017). "Contacts and meetings: Location, duration and distance traveled." In: *Travel Behaviour and Society* 6, pp. 64–74.

Tizzoni, Michele, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza (2014). "On the use of human mobility proxies for modeling epidemics." In: *PLoS computational biology* 10.7, e1003716.

Toch, Eran, Boaz Lerner, Eyal Ben-Zion, and Irad Ben-Gal (2018). "Analyzing large-scale human mobility data: a survey of machine learning methods and applications." In: *Knowledge and Information Systems*, pp. 1–23.

Tompkins, Adrian M, Laragh Larsen, Nicky McCreesh, and David Taylor (2016). "To what extent does climate explain variations in reported malaria cases in early 20th century Uganda?" In: *Geospatial health* 11.1s.

Toole, Jameson L, Yves-Alexandre de Montjoye, Marta C Gonzalez, and Alex (Sandy) Pentland (2015). "Modeling and Understanding Intrinsic Characteristics of Human Mobility." In: *Social phenomena : from data analysis to models*. Ed. by Bruno Gonçalves and Nicola Perrera. Springer. Chap. 2, pp. 15–35.

Townes, Lindsay R, Dyson Mwandama, Don P Mathanga, and Mark L Wilson (2013). "Elevated dry-season malaria prevalence associated with fine-scale spatial

patterns of environmental risk: a case–control study of children in rural Malawi." In: *Malaria journal* 12.1, p. 1.

Truscott, James and Neil M Ferguson (2012). "Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling." In: *PLoS Comput Biol* 8.10, e1002699.

Tsutsumi, Morito and Kazuki Tamesue (2012). "Intraregional Flow Problem in Spatial Econometric Model for Origin—Destination Flows." In: *Environment and Planning B: Planning and Design* 39.6, pp. 1006–1015.

Twitter (2018). URL: https://twitter.com (visited on 09/03/2018).

UK Data Service (2018). URL: http://wicid.ukdataservice.ac.uk/ (visited on 08/11/2018).

UN (2018). *2018 Revision of World Urbanization Prospects*. Tech. rep. UN:Department of social and economic affairs.

Valdez, LD, HH Aragão Rêgo, HE Stanley, and LA Braunstein (2015). "Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies." In: *Scientific reports* 5.

Van Audenhove, Francois-Joseph, Rominger Guillaume, Alexei Korn, Aurelia Bettati, Nicholas Steylemans, Michael Zintel, Andrew Smith, and Sylvain Haon (2018). *The future of mobility 3.0: Reinventing mobility in the era of disruption and creativity*. Tech. rep. Arthur D Little future lab.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry (2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." In: *Statistics and Computing* 27.5, pp. 1413–1432.

Viboud, Cécile, Ottar N Bjørnstad, David L Smith, Lone Simonsen, Mark A Miller, and Bryan T Grenfell (2006). "Synchrony, waves, and spatial hierarchies in the spread of influenza." In: *science* 312.5772, pp. 447–451.

Wai, Khin Thet, Myat Phone Kyaw, Tin Oo, PeThet Zaw, Myat Htut Nyunt, Moe Thida, and Thar Tun Kyaw (2014). "Spatial distribution, work patterns, and perception towards malaria interventions among temporary mobile/migrant workers in artemisinin resistance containment zone." In: *BMC public health* 14.1, p. 1.

Wall, Melanie M (2004). "A close look at the spatial structure implied by the CAR and SAR models." In: *Journal of Statistical Planning and Inference* 121.2, pp. 311–324.

Waller, Lance A, Bradley P Carlin, Hong Xia, and Alan E Gelfand (1997). "Hierarchical spatio-temporal mapping of disease rates." In: *Journal of the American Statistical association* 92.438, pp. 607–617.

Wallinga, Jacco and Peter Teunis (2004). "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures." In: *American Journal of Epidemiology* 160.6, pp. 509–516.

Weibo (2018). URL: https://weibo.com (visited on 09/03/2018).

Wesolowski, Amy, Caroline O Buckee, Linus Bengtsson, Erik Wetter, Xin Lu, and Andrew J Tatem (2014a). "Commentary: Containing the Ebola outbreak–the potential and challenge of mobile network data." In: *PLOS currents outbreaks*.

Wesolowski, Amy, Caroline O Buckee, Kenth Engø-Monsen, and CJE Metcalf (2016). "Connecting mobility to infectious diseases: the promise and limits of mobile phone data." In: *The Journal of infectious diseases* 214.suppl_4, S414–S420.

Wesolowski, Amy, Nathan Eagle, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee (2013). "The impact of biases in mobile phone ownership on estimates of human mobility." In: *Journal of the Royal Society Interface* 10.81, p. 20120986.

Wesolowski, Amy, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee (2012). "Quantifying the impact of human mobility on malaria." In: *Science* 338.6104, pp. 267–270.

Wesolowski, Amy, CJE Metcalf, Nathan Eagle, Janeth Kombich, Bryan T Grenfell, Ottar N Bjørnstad, Justin Lessler, Andrew J Tatem, and Caroline O Buckee (2015a). "Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data." In: *Proceedings of the National Academy of Sciences* 112.35, pp. 11114–11119.

Wesolowski, Amy, Wendy Prudhomme O'Meara, Nathan Eagle, Andrew J Tatem, and Caroline O Buckee (2015b). "Evaluating spatial interaction models for regional mobility in Sub-Saharan Africa." In: *PLOS Comput Biol* 11.7, e1004267.

Wesolowski, Amy, Taimur Qureshi, Maciej F Boni, Pål Roe Sundsøy, Michael A Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O Buckee (2015c). "Impact of human mobility on the emergence of dengue epidemics in Pakistan." In: *Proceedings of the National Academy of Sciences* 112.38, pp. 11887–11892.

Wesolowski, Amy, Gillian Stresman, Nathan Eagle, Jennifer Stevenson, Chrispin Owaga, Elizabeth Marube, Teun Bousema, Christopher Drakeley, Jonathan Cox, and Caroline O Buckee (2014b). "Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones." In: *Scientific reports* 4.

White, Laura F, Brett Archer, and Marcello Pagano (2013). "Estimating the reproductive number in the presence of spatial heterogeneity of transmission patterns." In: *Int J Health Geogr* 12, p. 35.

White, Laura F, Brett Archer, and Marcello Pagano (2014). "Determining the dynamics of influenza transmission by age." In: *Emerging themes in epidemiology* 11.1, p. 4.

White, Laura Forsberg and Marcello Pagano (2008). "Transmissibility of the influenza virus in the 1918 pandemic." In: *PLoS One* 3.1, e1498.

WHO (2017). URL: http://apps.who.int/gho/data/node.ebola-sitrep.quick-downloads?lang=en (visited on 12/12/2017).

WHO, World Health Oganization Ebola Response Team (2014). "Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections." In: *N Engl J Med* 2014.371, pp. 1481–1495.

WHO, World Health Organization Ebola Response Team (2016). "After Ebola in West Africa—unpredictable risks, preventable epidemics." In: *New England Journal of Medicine* 375.6, pp. 587–596.

Williams, Nathalie E, Timothy A Thomas, Matthew Dunbar, Nathan Eagle, and Adrian Dobra (2015). "Measures of human mobility using mobile phone records enhanced with GIS data." In: *PloS one* 10.7, e0133630.

Wilson, Alan G (1967). "A statistical theory of spatial distribution models." In: *Transportation research* 1.3, pp. 253–269.

Wood, CL, A McInturff, HS Young, D Kim, and KD Lafferty (2017). "Human infectious disease burdens decrease with urbanization but not with biodiversity." In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372.1722.

World Health Organization(WHO) (2015). *World malaria report 2015*.

World Pop (2018). URL: http://www.worldpop.org.uk/ (visited on 09/16/2018).

Wu, Xiaoxu, Yongmei Lu, Sen Zhou, Lifan Chen, and Bing Xu (2016). "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation." In: *Environment international* 86, pp. 14–23.

Xia, Yingcun, Ottar N Bjørnstad, and Bryan T Grenfell (2004). "Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics." In: *The American Naturalist* 164.2, pp. 267–281.

Yadav, Kavita, Sunil Dhiman, Bipul Rabha, PK Saikia, and Vijay Veer (2014). "Socio-economic determinants for malaria transmission risk in an endemic primary health centre in Assam, India." In: *Infectious diseases of poverty* 3.1, p. 19.

Yan, Xiao-Yong, Chen Zhao, Ying Fan, Zengru Di, and Wen-Xu Wang (2014). "Universal predictability of mobility patterns in cities." In: *Journal of The Royal Society Interface* 11.100, p. 20140834.

Yang, Fen, Lingling Yuan, Xuhui Tan, Cunrui Huang, and Jun Feng (2013). "Bayesian estimation of the effective reproduction number for pandemic in-

fluenza A H1N1 in Guangdong Province, China." In: *Annals of epidemiology* 23.6, pp. 301–306.

Yang, Yingxiang, Carlos Herrera, Nathan Eagle, and Marta C Gonza (2014a). "Limits of Predictability in Commuting Flows in the Absence of Data for Calibration." In: *Scientific reports* 4, pp. 1–9.

Yang, Yingxiang, Carlos Herrera, Nathan Eagle, and Marta C González (2014b). "Limits of predictability in commuting flows in the absence of data for calibration." In: *arXiv preprint arXiv:1407.6256*.

Yue, Ricci PH, Harry F Lee, and Connor YH Wu (2017). "Trade routes and plague transmission in pre-industrial Europe." In: *Scientific reports* 7.1, p. 12973.

Zhang, Ying, Peng Bi, and Janet E Hiller (2008). "Climate change and the transmission of vector-borne diseases: a review." In: *Asia-Pacific Journal of Public Health* 20.1, pp. 64–76.

Zhao, Ziliang, Shih-Lung Shaw, Yang Xu, Feng Lu, Jie Chen, and Ling Yin (2016). "Understanding the bias of call detail records in human mobility research." In: *International Journal of Geographical Information Science* 30.9, pp. 1738–1762.

Zheng, Yan-Tao, Zheng-Jun Zha, and Tat-Seng Chua (2012). "Mining travel patterns from geotagged photos." In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 3.3, p. 56.

Appendices

Appendix A

Malaria mapping in Malawi

The expected malaria case-counts e_{st} were calculated by multiplying the overall malaria risk for Malawi, π and the population of each district p_{st} , i.e. $e_{st} = p_{st}\pi$. The overall malaria risk for Malawi is given by dividing the total number of cases in Malawi with the total population, i,e, $\pi = \frac{\sum y_{st}}{\sum p_{st}}$ where y_{st} is the observed malaria counts in district s at time t and p_{st} is the corresponding population. The logarithm of the expected malaria counts is then included in the model as offset with a coefficient of 1 and hence no effect on the response variable. The standardised morbidity ratio (SMR) is given by the ratio of observed estimated cases, i.e. $SMR = \frac{y_{st}}{e_{st}}$. The maximum likelihood estimate (MLE) of the risk R_{st} of a district in a GLM without random effects is the corresponding SMR, i.e. $R_{st} = \frac{y_{st}}{e_{st}}$. In the mixed model setting, the posterior mean of the relative risk for each district is therefore a weighted average of the SMR for the district and the prior mean of the relative risk in the overall spatial region giving rise to smoother estimates than those provided by raw SMR estimates.

Appendix B

Model diagnostics

Figure B1 is the traceplot showing convergence.

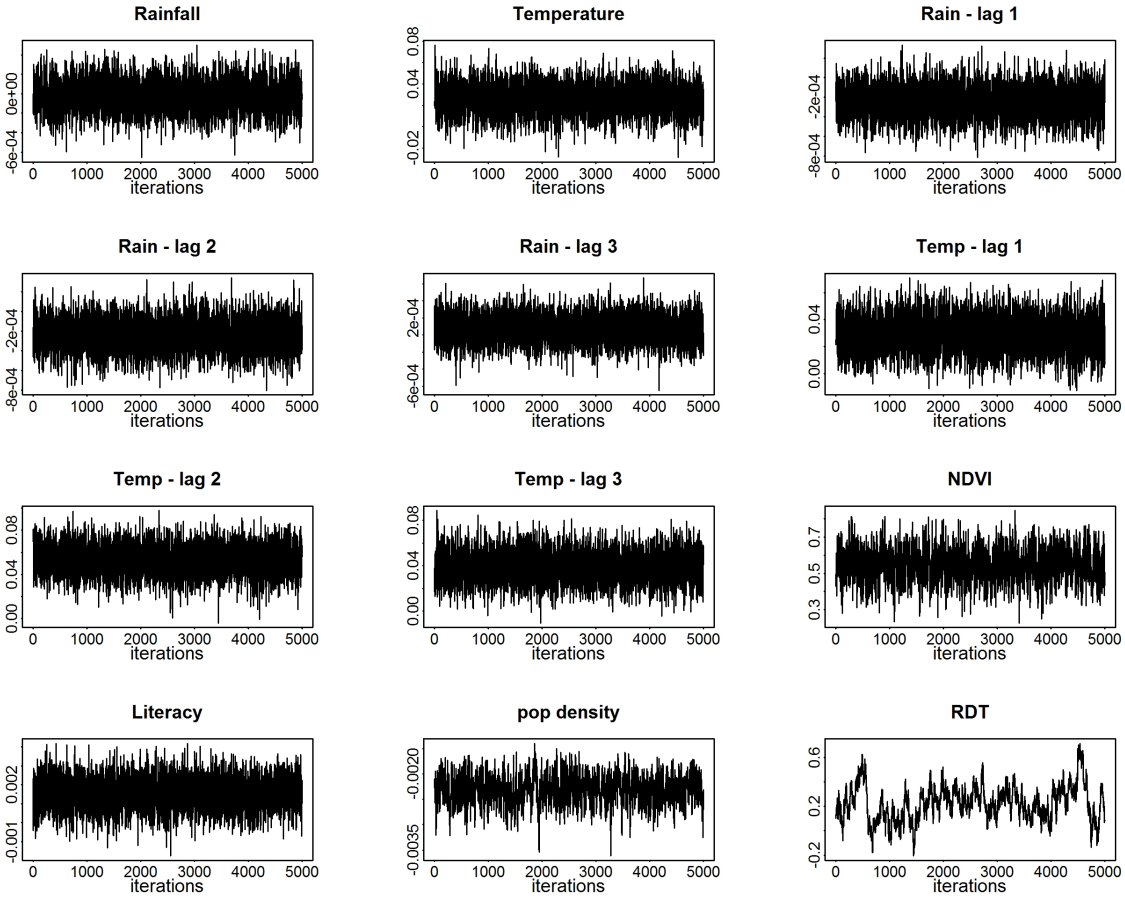

Figure B1: Trace plots for the covariates in the final model

Appendix C

Epidemic curves

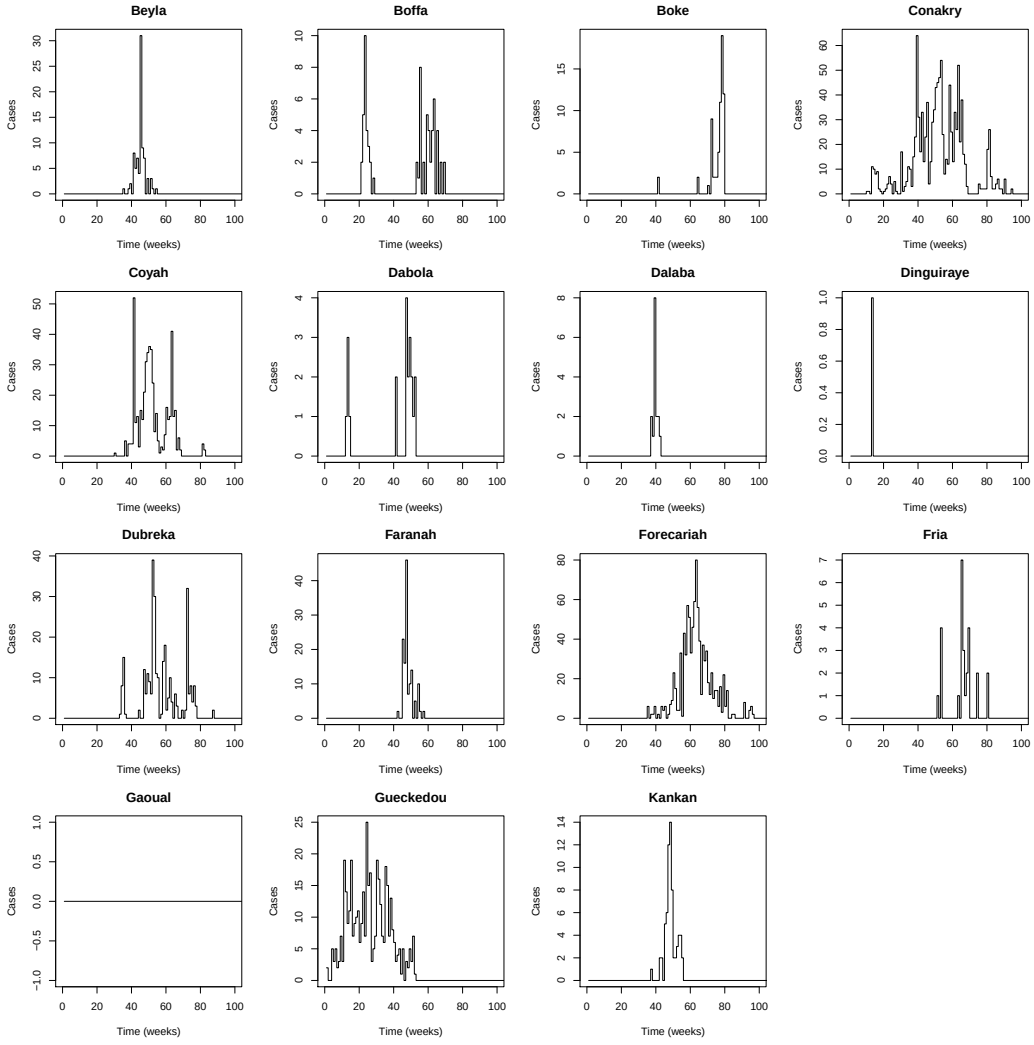

Figure C1: Weekly epidemic curves for the districts in Guinea

The epidemic curves for Liberia are shown in figure

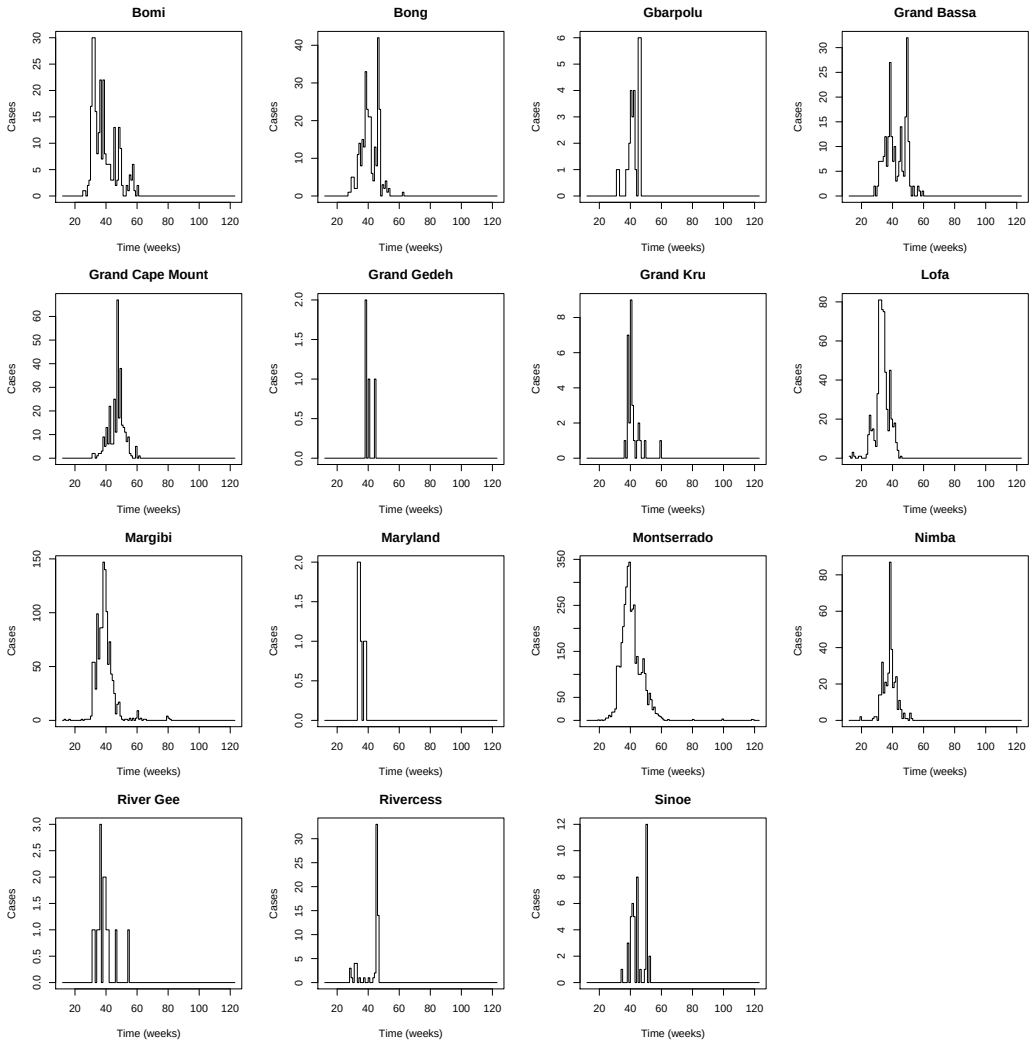

Figure C2: Weekly epidemic curves for districts in Liberia

For Sierra Leone, the epidemic curves are

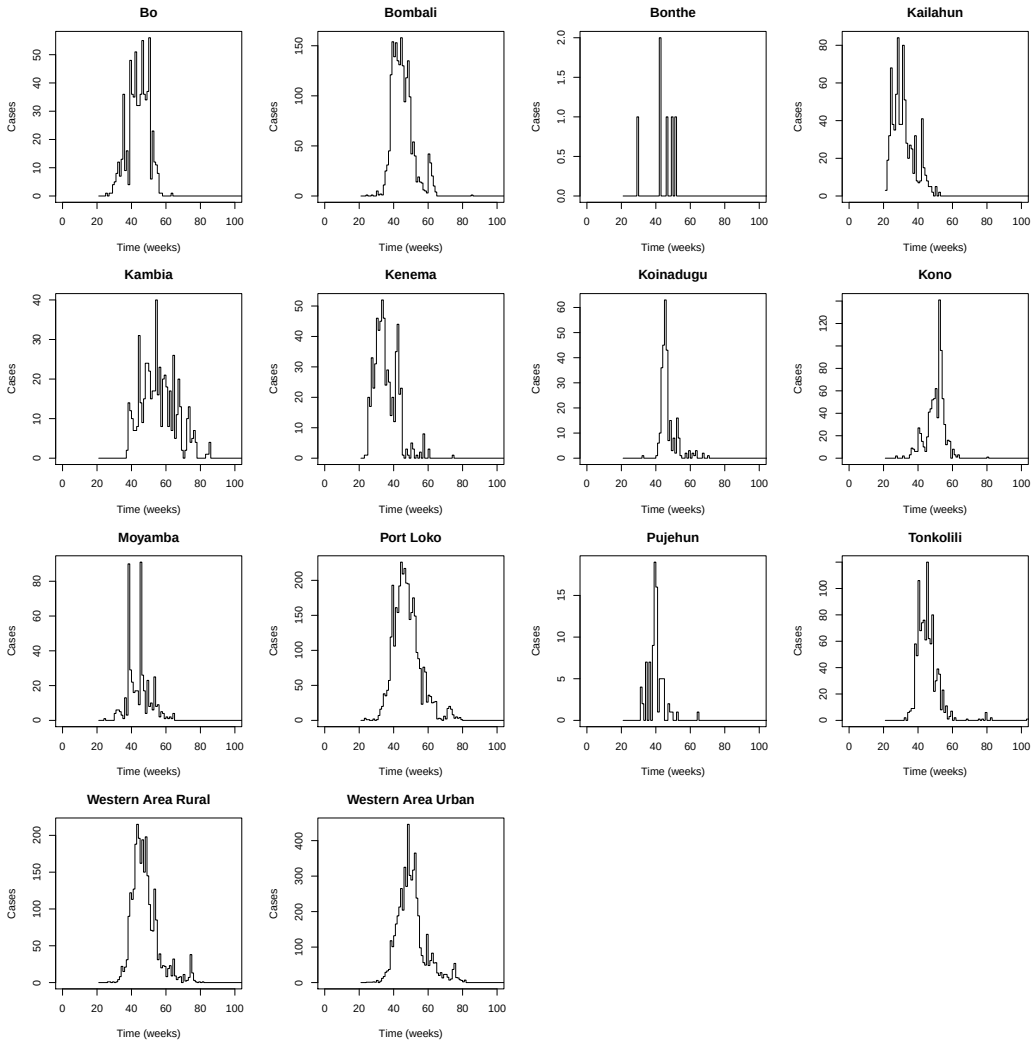

Figure C3: Weekly epidemic curves for Sierra Leone districts

Appendix D

Spatial reproductive number estimates

In this section, we present the estimates for Guinea and Sierra Leone

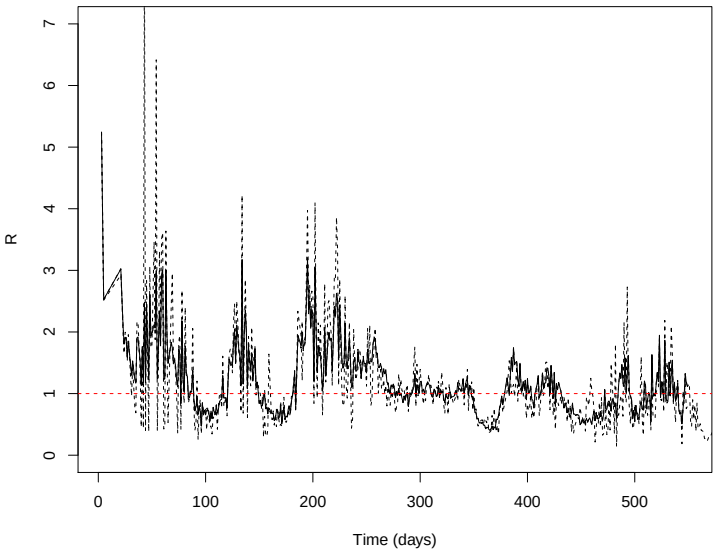

Figure D1: Overall estimate of the the spatial reproductive number $R(x, y, t)$ for Guinea

And for Sierra Leone

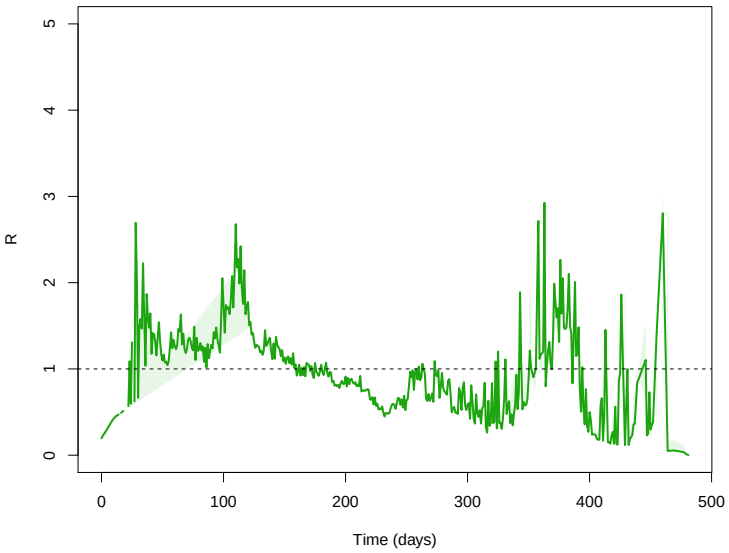

Figure D2: Overall estimate of the the spatial reproductive number $R(x, y, t)$ for Sierra Leone

Within spatial unit estimates for Guinea

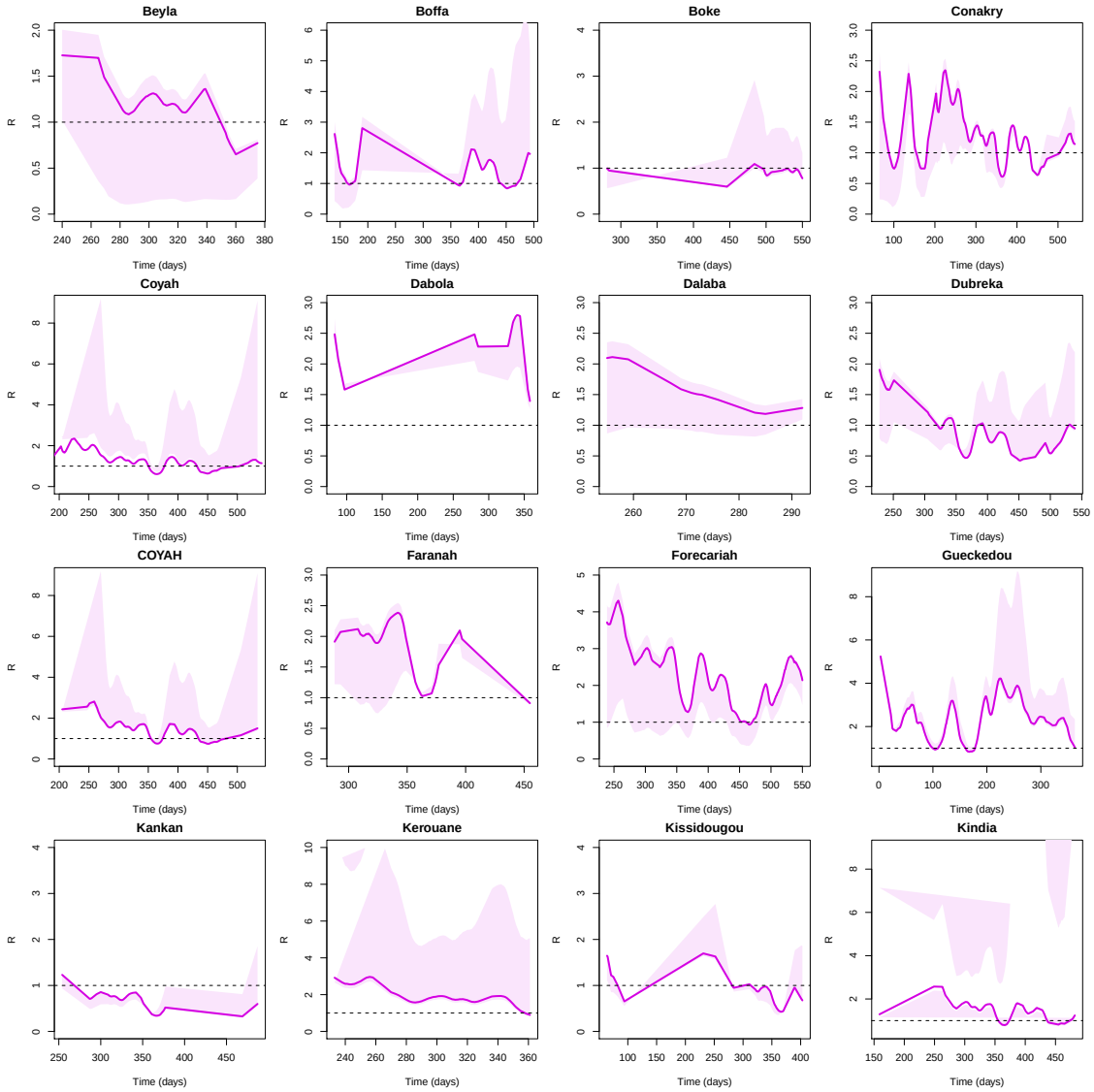

Figure D3: Spatial temporal estimates $R(y, y, t)$ within districts in Guinea

For Guinea,WT estimates is given in the figure

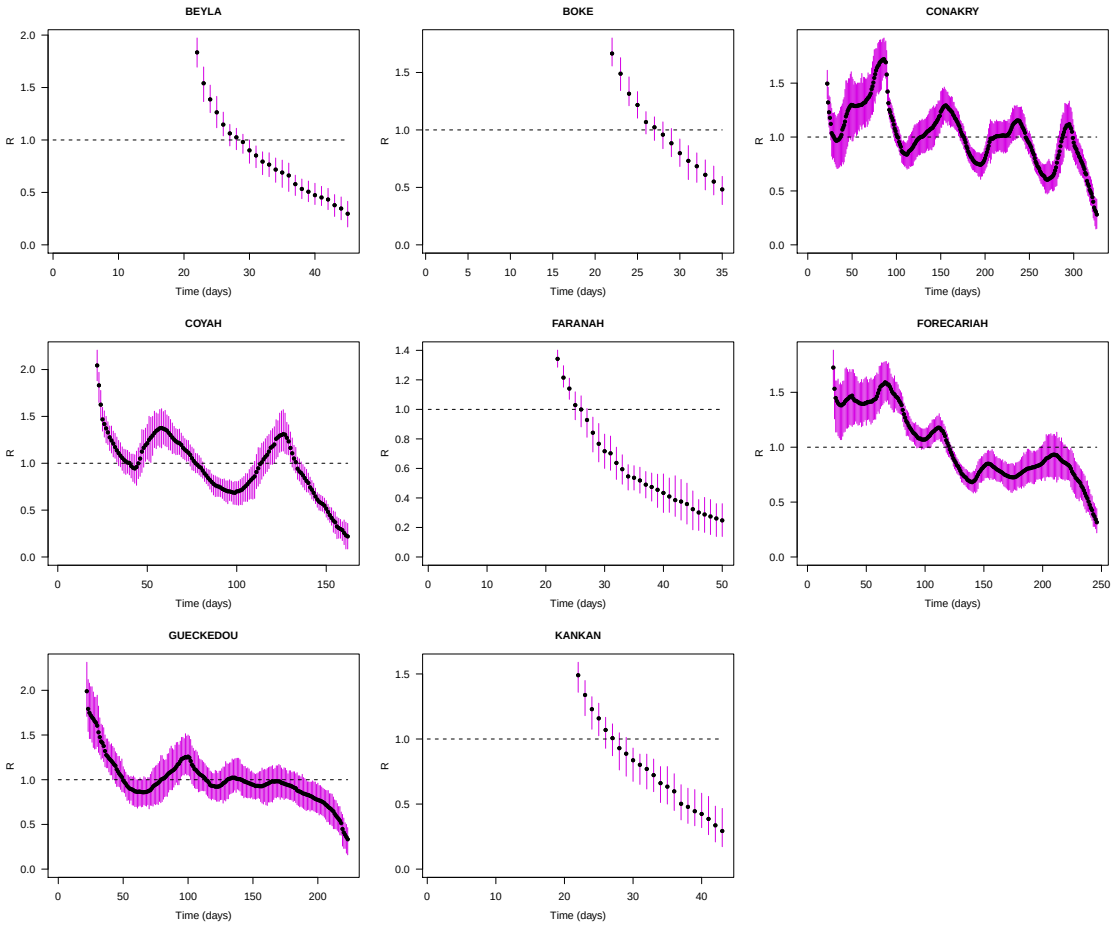

Figure D4: WT estimates for selected districts in Guinea

The spatial R_t for Sierra Leone are shown in the figure D5

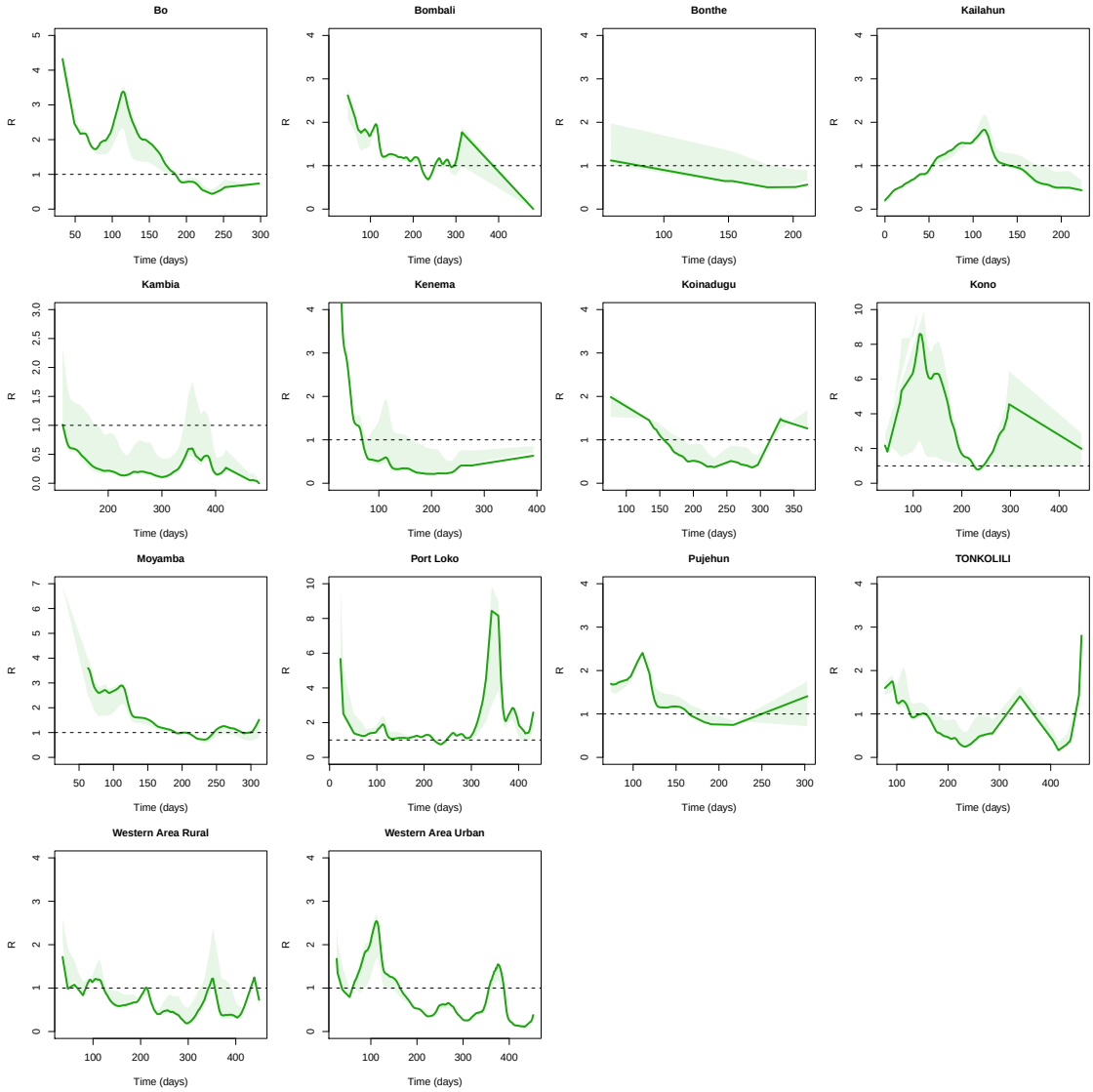

Figure D5: Spatial temporal estimates $R(y, y, t)$ within districts in Sierra Leone

The Wallinga Tuenis estimates for the districts in Sieraa Leone are given in figure D6

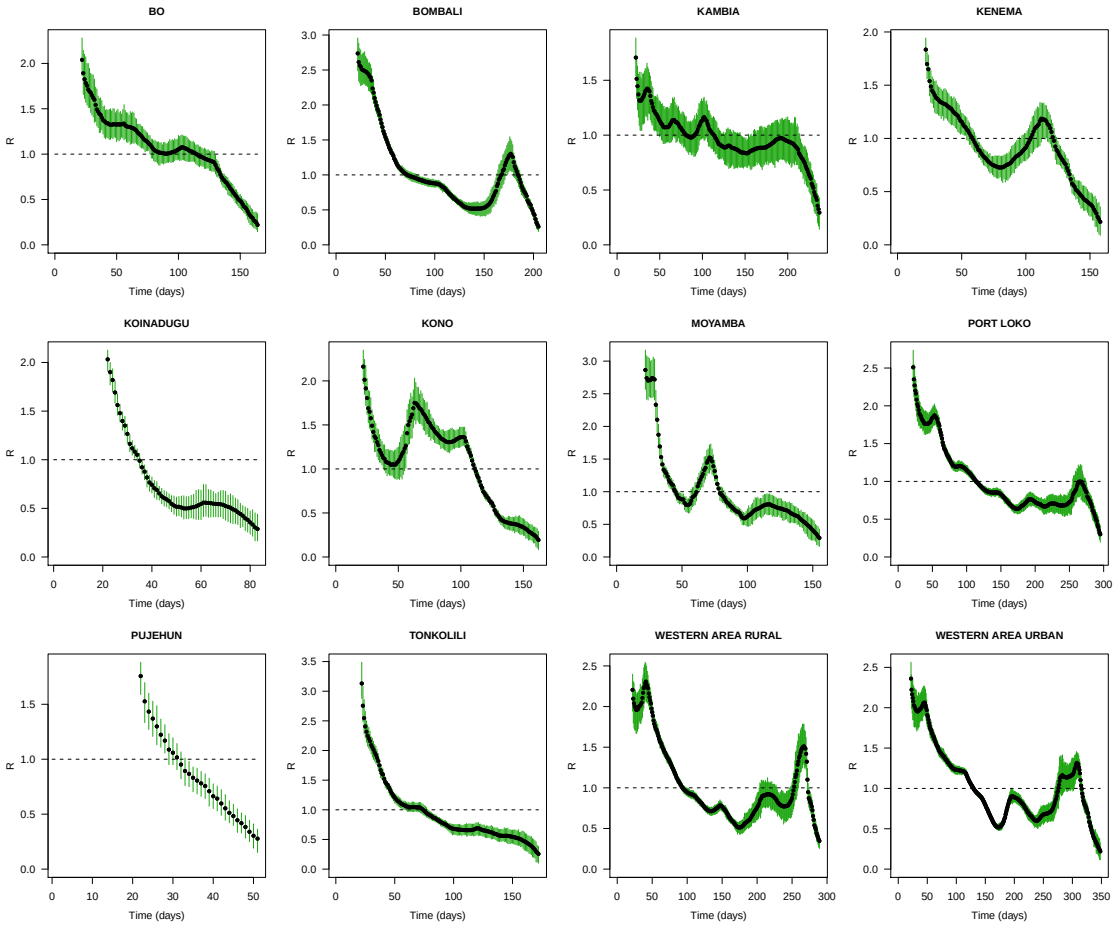

Figure D6: WT estimates for districts in Sierra Leone