MODELLING SPATIO-TEMPORAL PATTERNS OF DISEASE RISK FOR DATA WITH MISALIGNMENT AND MEASUREMENT ERRORS: AN APPLICATION ON MEASLES AND HIV PREVALENCE DATA IN NAMIBIA

A DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN SCIENCE (STATISTICS)

OF

THE UNIVERSITY OF NAMIBIA

BY

DISMAS NTIRAMPEBA

9711058

APRIL 2018

MAIN SUPERVISOR: PROF LAWRENCE KAZEMBE

COSUPERVISOR: DR ISAAK NEEMA

Abstract

Disease mapping has important applications in public health because it enables the identification of areas which are at high risk of particular health problems. It helps visualising the spatial pattern of the disease distribution, which is of interest to the health sector as it enables the sector to plan, evaluate and redesign prevention and control strategies, and also make important policy decisions particularly for geographically targeted intervention in resource poor settings. Analyses of spatial disease patterns are generally based on data of a single disease and they are often fraught with challenges that include lack of a representative sample, often incomplete and most of which may have measurement errors, and may be spatially and temporally misaligned. This thesis focused on the development and extension of statistical models with particular interest to dealing with misalignment, measurement errors and jointly modeling of data from multiple sources.

The first objective was to estimate and map the risk of measles at a sub-region level (i.e. constituency level) in Namibia using data obtained at the regional level. Direct inferences at constituency level made on basis of the original level of aggregation may lead to an inferential problem known as a misalignment in the statistical literature. Using measles data from Namibia for the period 2005-2014, both multi-step and direct approaches were applied to correct the misalignment. The multi-step approach model provided a relatively better model.

The second objective was to fit a spatio-temporal model while dealing with misalignment and measurement error, again applied to measles data aggregated at regional level over the period 2005 to 2014. Again this leads to a spatial misalignment problem if the purpose is to make decisions at constituency level. Moreover, data on risk factors of measles were not available each year between 2005 and 2014. Thus, assuming that covariates were constant through the study period would induce measurement errors which might have effects on the analysis results. The multi-step approach was further extended to include temporal effects and account for measurement errors. Consequently, spatio-temporal models, which included Bernardinelli and Knorr-Held approaches, and classical measurement error models were adopted. Comparison of the results obtained from the naïve method (i.e. modelling that

ignored errors in covariates) and those from the approach that accounts for measurement errors showed that the latter modelling approach performed better than the former. The study showed a spatio-temporal variation of the measles risk over the 2009-2014 period.
The third objective of this study was to develop a joint spatial model for HIV prevalence, using two sources (i.e. 2014 National HIV Sentinel survey (NHSS) among pregnant women aged 15-49 years attending antenatal care (ANC) and the 2013 Namibia Demographic and Health Surveys (NDHS)), which would enable the estimation at any location of the constituency or district level while dealing with misalignment in data. The shared component modelling approach was adopted through the use of stochastic partial differential equations (SPDE). The bivariate modelling approach developed allowed to combine two data sources that are available at different spatial levels in a single model and it catered for a specification of different spatial processes through the link function. Findings revealed that health districts and constituencies in the northern part of Namibia were highly associated with HIV infection. Also, the study showed that the place of residence, gender, gravida, marital status, number of kids dead, wealth index, education, and condom use were significantly associated with HIV infection in Namibia. Finally, it was shown that the prediction of HIV prevalence using the NDHS data source can be enhanced by jointly modelling other HIV data such as NHSS data.

In conclusion, results showed that the multi-step approach may be used to deal with misalignment. Moreover, introducing the error model proved to be a useful approach to correct for measurement errors in data and improve inferences in situations where mismeasured values in covariates are encountered instead of naïve analyses that ignore the presence of errors in measurements. Lastly, the thesis showed that the prediction of HIV prevalence using the NDHS data source can be enhanced by jointly modelling other HIV data such as NHSS data.

List of publications and conference presentations

1. Ntirampeba, D., Neema, I., & Kazembe, L. (2017). Modelling spatial patterns of misaligned disease data: An application on measles incidence in Namibia. *Clin Epidemiol Glob Health*. doi.org/10.1016/j.cegh.2017.01.002.

2. Ntirampeba, D., Neema, I., & Kazembe, L. (2017). Joint spatial modelling of disease risk using multiple sources: An application on HIV prevalence from antenatal sentinel and demographic and health surveys in Namibia. *Global Health Research and Policy*, 2017, 2:22. doi 10.1186/s41256-017-0041-z.

3. Ntirampeba, D., Neema, I., & Kazembe, L. (2017). Modelling spatio-temporal patterns of disease for spatially misaligned data: An application on measles incidence data in Namibia, 2005-2014. *Submitted*.

4. Modelling spatio-temporal patterns of disease for spatially misaligned data: An application on measles incidence data in Namibia, 2005-2014. *Presented at UNAM 2nd Multi/Interdisciplinary Research Conference.*

Acknowledgements

I would like to thank my supervisors, Prof Lawrence Kazembe and Dr Isaak Neema, for their brilliant ideas, guidance, support, friendship, and dedication to the work throughout the duration of my PhD. Both of them have helped make my experience an almost entirely positive one. Without them, this work would never have been completed.

I would also like to thank my late parents without whom I would never have been here to complete this work. Thanks to my wife, Ngomirakiza, my daughter, Kaze and my son Van Fundi, for their constant support, their understanding during the more difficult spells, and for always keeping me positive and cheerful.

Sincere gratitude is extended to Charlotte Jones-Todd for her assistance in R programming. Thanks to my colleagues at Namibia University of Science and Technology and friends who supported my ambition and walked with me in this academic journey. In particular, I would like to thank Prof Jean Baptiste Gatsinzi for his guidance in this dissertation write up.

Lastly but not least, I wish to thank the Almighty God for giving me strength, accompanying me this far, and granting me His endless mercies.

Declarations

I, Dismas Ntirampeba, hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

No part of this thesis/dissertation may be reproduced, stored in any retrieval system, or transmitted in any form, or by means (e.g. electronic, mechanical, photocopying, recording or otherwise) without the prior permission of the author, or The University of Namibia in that behalf.

I, Dismas Ntirampeba, grant The University of Namibia the right to reproduce this thesis in whole or in part, in any manner or format, which The University of Namibia may deem fit.

....

Name of Student Signature Date

Contents

List of Tables

List of Figures

List of Abbreviations and/ or Acronyms

AIDS	Acquired Immunodeficiency Syndrome
ANC	Antenatal Clinic
BYM	Besag -York - Mollie
CAR	Conditional Autoregressive
CI	Credible Interval
COSP	Change of Support Problem
DIC	Deviance Information Criterion
EA	Enumeration area
GIS	Geographic Information System
GLS	Generalised Least Squares
GMRF	Gaussian Markov Random Field
GRF	Gaussian Random Field
HIV	Human Immunodeficiency Virus
HMIS	Health Management Information System
ICAR	Intrinsic Conditional Autoregressive
IG	Inverse Gamma
IGLS	Iterative Generalised Least Squares
IID	Independently Identically Distributed
INLA	Integrated Nested Laplace Approximation
MACR	Multivariate Conditional Autoregressive
MAUP	Modifiable Areal Unit Problem
MCMC	Markov Chain Monte Carlo
MDG	Millennium Development Goals
MH	Metropolis-Hastings
MoHSS	Ministry of Health and Social Services
MSRE	Mean Root Square Error
MVN	Multivariate Normal
N	Normal (distribution)
NB	Negative Binomial
NDHS	Namibia Demographic and Health Survey
NIP	Namibia Institute of Pathology

NHSS	National HIV Sentinel Surveillance
NPC	National Planning Commission
NPHC	Namibia Population and Housing Census
NSA	Namibia Statistics Agency
OLS	Ordinary Least Squares
OR	Odds Ratio
PAHO	Pan American Health Organisation
PLS	Partial Least Squares
PMTC	Prevention of Mother To Child Transmission
RAMPS	Reparameterised And Marginalised Posterior Sampling
SAR	Simultaneous Autoregressive
SDG	Sustainable Development Goal
SRR	Specific Relative Risk
SPDE	Stochastic Partial differential Equation
STI	Sexually Transmitted Infection
U	Uniform (distribution)
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization
ZINB	Zero Inflated Negative Binomial
ZIP	Zero Inflated Poisson

Chapter 1

Introduction

1.1 An overview

Recent advances in geographic information systems (GIS) and the internet make it possible to access spatial data in various forms that include point, line, area, surface, etc. Most of the data are collected using surveys and they are spatially referenced by locations that include districts, regions, constituencies, or any other administrative areal units. With the increasing availability of geographically referenced data, linking of collected data is indeed inevitable as the exploitation of this readily available information helps in avoiding the implementation of new and expensive data collection methods. But one major concern is how best these data can be integrated to answer real life problems. The integration of such information may require the data transformation as the spatial process of interest intrinsically present in one form of data may be completely different from the one observed in another form of data. Thus, the development and application of spatial methods have to deal with the issue of misalignment. In addition to misalignment, data might have errors.

1.2 Conceptual framework and background to spatial modelling

1.2.1 Conceptual framework

An interesting question which is quite common in public health studies is to find out if people of similar characteristics will experience different health outcomes when located in different areas. A positive answer will then imply that the area promotes or inhibits health, over and above individual socio-economic characteristics. In their study on neighborhoods and health, Diez Roux & Mair (2010) re-emphasized the existence of the effects of physical and social neighborhood environments on the health of residents of any communities or locations. They further indicated that a better understanding of health or disease distributions requires both individual and group characteristics.

Macintyre, Ellaway, & Cummins (2002) suggested three explanations for geographical variations in health. First, the characteristics of individuals concentrated in particular places also known as compositional explanations; second, opportunity structures in physical environments (for example, the availability of health environments and public or private services in support of people in their daily lives); and third, socio-cultural and historical features of communities. Macintyre et al. (2002) further linked these three explanations to Maslow hierarchy of human needs that range from air to social, cultural, and physical recreation needs. Macintyre et al. (2002) finally suggested that the basis for discovering how places impact on health is to make use of a framework of universal human needs.

This study is hinged on such concepts as its conceptual framework to explain specific pathways by which an area may affect, directly or indirectly, the spatial distribution of disease risk. Furthermore, the study followed the ontological framework of space and time models by Peuquet (1999) expressed in a triad conceptual model (i.e. what/where/when model) illustrated in Fig. 1.1. This concept allowed for the inclusion of space, time, and interactions of space in modelling the distribution of diseases burden (i.e measles and HIV).

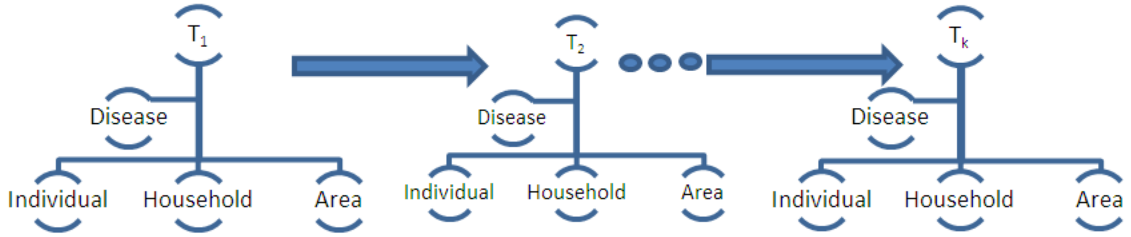

Figure 1.1: Illustration of a triad model: what/where/when

1.2.2 Background to spatial modelling

The classical book by Cressie (1993) distinguishes three main inferential frameworks in spatial analysis, namely spatial point pattern analysis, geostatistical analysis, and lattice or areal statistical analysis. In each of these inferential approaches, the Bayesian inferential framework is preferred as it permits to derive posterior predictive distributions for both parameters and epidemiological outcomes of interest. It is also suitable when dealing with multiple sources of uncertainty and it enables to incorporate additional sources of information in the form of prior knowledge (Liang, Banerjee, Bushhouse, Finley, & Carlin, 2008). For areal models, conditional autoregressive (CAR) pioneered by Besag (1974) and further developed by Besag, York, & Mollie (1991) has been commonly used to account for spatial variation (Jin, Carlin, & Banerjee, 2005)).

In spatial analysis, there exist many challenges that include joint the modelling of data from multiple sources, measurement errors in covariates or response variable, specification of spatial-temporal trends, big "N" problem, "knotty" problem, and boundary problem. In this review, the focus is on the following three main challenges that constitute the core of the proposed study. The first challenge is to combine data from multiple sources of data; followed by the measurement errors; and the third issue is the specification of spatial-temporal trends.

Combining data from multiple data sources has several advantages. For example, it may be expensive to conduct a new study for every new problem of interest. Thus, combining information from various sources can enable one to obtain more information in the face of limited resources (Schenker, 2013). Also, merging information from different sources helps to improve the estimation of related measures as one source may be used to provide missing information in another (He, Landrum, & Zaslavsky, 2014). Moreover, extracting information on the same collection of variables but reported from different sources may lessen non-sampling errors that include coverage error, errors due to missing data, nonresponse errors, and measurement errors (Raghunathan et al., 2007). As a result of these strengths, a combination of information from different sources is commonly used. For instance, Schenker, Gentleman, Rose, Hing, & Shimizu (2002) combined data from two surveys in order to extend the coverage; Raghunathan et al. (2007) pooled together data from two surveys to estimate cancer prevalence rates; Schenker, Raghunathan, & Bondarenko (2010) used information from an examination-based survey to enhance the analysis of self-reported data in a large-scale health survey; and Manda, Masenyetse, Cai, & Meyer (2015) jointly analysed the data from NDHS and ANC surveys to map HIV prevalence. However, combining data might lead to problems if approriate methods are not employed. For instance, if areal and geostatistical data are available, one natural and commonly used way to combine data is to aggregate the geostatistical data. Specific examples are naïve kriging methods and area-to-point kriging methods to collapse data in geographical units into their centroids (Goovaerts, 2008). But by doing so, one loses attributes at individual level. Merging together data sources available at different levels of aggregation may suffer from a misalignment problem.

As much as the use of multiple source data may lead to bias in estimates due to misalignment, the measurement error in spatial analysis is also a recognized serious influential factor of wrong inferences of spatial effects. Measurement error occurs in almost every discipline. Scholars have warned that ignoring measurement error may for example lead to masking some important features of data, losing the power of hypothesis testing of relationships among variables, and introducing bias in estimates (Wattanasaruch, Pongsapukdee, & Khawsithiwong, 2012). Some of the works dedicated to this problem include accounting for measurement error in covariates through the use of hierarchical modelling in disease mapping (Xia & Carlin, 1998); use of a structural modelling approach to deal with spatial covariates measured with errors (Yi, Tang, & Lin, 2009); adjustment of prevalence estimates using sensitivity and specificity (Njai, Siegel, Miller, & Liao, 2011); computing adjusted prevalence of transmitted HIV drug resistance as a ratio between a function of experienced tests with resistance and a function of total naïve tests (Castro, Pillay, Sabin, & Dunn, 2012); estimation of a function of unobserved true values using non-parametric deconvolution techniques, and the correction of measurement error and /or misclassification using calibration methods (Wattanasaruch et al., 2012); and correcting bias in ischemic heart disease using a semi-parametric regression approach (Huque, Bondell, Carroll, & Ryan, 2016).

Apart from the challenges arising from misalignment and measurement error, the use of crude data to construct disease maps may lead to inaccurate maps (Hampton, Serre, Gesink, Pilcher, & Miller, 2011). Statistical analysis results are suitable for the construction of accurate disease maps. Various techniques, which include among others kriging, generalized linear and generalized linear mixed effects models are commonly used (Diggle, Tawn, & Moyeed, 1998; Hampton, Serre, Gesink, Pilcher, & Miller, 2011; Lentz, Blackburn, & Curtis, 2011). Further examples include zero inflated Poisson (ZIP) models which have been used to map spatial relative risks meningococcal disease across Germany in 2004 (Gschlößl & Czado, 2008) and diabetes incidence distribution patterns in the youth of South Caroline (Song, Lawson, Agostino, & Liese, 2011); Mohebbi, Wolfe, & Forbes (2014) employed negative binomial and Poisson disease for mapping esophageal cancer incidence data in the Caspian region of Iran; Arab (2015) analysed the spatio-temoral distribution of Lyme disease in the Illinois county using hurdle models (Poisson and negative binomial) and zero-inflated models (Zero inflated Poission and Zero inflated negative binomial(ZINB)); and Neyens et al. (2017) investigated the distribution of mesothelioma in Flanders through Bayesian disease mapping models (i.e. combined Poisson model and combined hurdle model).

All these models can be generally extended to spatio-temporal models by including a temporal component in the model. A review of some of the early work done by different researchers in the context of count data modeling can be found in Lawson (2013).

1.3 Orientation of the study and statement of the research problem

1.3.1 Orientation of the study

Worldwide, there is a general increase in the burden caused by expenses related to Human Immunodeficiency Virus (HIV) and related diseases or other diseases for most countries. For example, UNAIDS World AIDS Day Report 2012 indicates that domestic investments for the AIDS response increased from 3.9 billion US dollars in 2005 to almost 8.6 billion US dollars in 2011.

In Namibia, an HIV prevalence of 13.4% was observed in 2010/11 (UNAIDS, 2013). Despite the encouraging news that about 70,000 new HIV infections are being avoided annually, it is estimated that 29.7% of pregnant women aged between 35 years and 39 years are HIV positive (UNAIDS, 2013). As a process towards a Namibia free of AIDS, the Namibian Government has increased domestic investments for the AIDS response. For instance, in the financial year 2008/09, the government funded 45.5% of the nation's HIV response, which represented over 28% of the total health budget (UNAIDS, 2013).

Furthermore, worldwide, measles is ranked among the leading causes of mortality especially among children in developing countries. For instance, in 2013, about 145,700 deaths were recorded (WHO, 2015). Deaths due to measles are quite common among malnourished children and people whose immune system has been weakened by diseases that include the human immunodeficiency virus/ acquired immunodeficiency syndrome (HIV/AIDS). High death rates are commonly registered in developing countries with low income per capita and poor health service systems (WHO, 2014). As there is no antiviral treatment for the measles virus so far, measles vaccination and supportive care that includes good nutrition and adequate fluid intake have been used to fight measles (WHO, 2015). But, the reduction of global funding has largely affected the immunisation campaigns, which has hampered the efforts of a complete elimination of measles (WHO, 2014). Consequently, measles cases are still reported in many countries, with Angola, Ethiopia, Namibia, Bosnia and Herzegovina, Georgia, Sri Lanka, and the Philippines being ranked among the top ten countries with high annualised measles incidence per 100 000 inhabitants in 2014 (WHO, 2017). Therefore, it is crucial to find ways to fight against HIV and measles and ensure that funds allocated to the fight against these diseases are used efficiently.

Maps are often used to spot out areas of a country with the most disease occurrences in order to plan for a proper intervention and targeted distribution of aid to most affected areas. They are indeed regarded as useful tools for geographical targeted interventions, monitoring and evaluation of disease burden. However, the construction of such maps is fraught with a number of challenges. One of the setbacks is that these maps are constructed using data that may contain errors. Also, these maps may inherit the problems due to biased selection methods and the sparse nature of data collected from small geographical areas or ecological fallacy.

For instance, in Namibia, as in many countries, HIV statistics from prevention of mother to child transmission (PMTCT), antenatal clinics and syphilis surveillance, and antiretroviral treatment programs are available at site or health district levels. The availability of many sources of HIV data provides an opportunity to use a combination of data sources in order to provide more precise estimates and hence produce maps with high resolution. However, the combination of different data sources has its own challenge as data may be spatio-temporally misaligned. Many scholars have almost exclusively opted for modelling data from different sources separately with antenatal care (ANC) clinics considered as the best possible source information on HIV.

In many physical phenomena that include those in epidemiological, ecological and environmental studies, it is common to encounter situations of having data aggregated at one level but the problem of interest requires making decision at a different level of aggregation. In particular, for disease surveillance, disease data are commonly available in aggregated format in order to preserve the privacy of patients. In Namibia, measles data were available in aggregated format at regional level over the period 2005 to 2014. Yet, health decisions might be needed at lower administrative boundaries such as constituency level. If direct inferences at constituency level are made on the basis of the regional data, then such inferences may suffer from the problem known as misalignment (Finley, Banerjee, & Cook, 2014). Although many methods of dealing with this type of misalignment are found in literature, most of them rely on the availability of covariates at both levels of aggregation. In addition to the misalignment issue, socio-economic variables related to measles were not measured for each year, rather surrogate variables were used instead. Assuming no measurement errors in these covariates would introduce bias in estimates and consequently lead to erroneous conclusions (Wattanasaruch et al., 2012; Buonaccorsi, 2010). Despite rich literature on measurement error modelling, many researchers still use naïve modelling approach that assumes that covariates are observed without errors.

Therefore, the purpose of this study was to use measles data and a combination of data obtained from different sources while adjusting for measurement errors in order to derive reliable small area estimates of measles and HIV risk that would enable the government and various policy and decision makers to deal with issues of the distribution of funds and resources, equity, disparity, intervention and surveillance programs. Ultimately, high resolution maps of diseases risks in Namibia were constructed.

1.3.2 Statement of the research problem

In recent years, maps depicting geographical distributions of diseases at small areas have been used extensively in public health, mostly in the analysis of disease risk distribution. An understanding of the magnitude and spatial and/or temporal distribution of disease risks is essential for planning, evaluating and re-designing prevention and control strategies, and other important policy decisions particularly for geographical target interventions in resource poor settings. In modeling the spatial distribution of disease burden, it is assumed that the analysis is based on best possible data. However, for instance HIV data are fraught with challenges; including lack of representative samples, mostly depending on sentinel data, often incomplete data and most of which may have measurement errors such as misclassification due to self-reporting, inaccurate measuring instruments, poor data coding and recording, and poor data management. Additionally, combining data from two sources, namely the 2014 National HIV Sentinel survey (NHSS) and the 2013 Namibia Demographic and Health Surveys (NDHS), would pose a misalignment as these two surveys are conducted at different levels of aggregation. For measles, health policy makers might be interested at inferring at lower level (i.e. constituency level), yet data is available at regional. Direct inferences at constituency level using data originally collected at regional level might be troubled by misalignment. Also, the use of surrogates of socio-economic variables related to measles in naïve models would induce bias in estimates as results of errors in covariates. All these problems pose a complex modeling challenge when analysing the risk of HIV as well as measles at small areas in many African countries, including Namibia. Nevertheless, an effective fight against HIV epidemic and measles requires the use of maps constructed with reliable risk estimates which constitute an alternative approach for monitoring and evaluation. To the best of our knowledge, detailed spatial analyses of HIV and measles data have not been done in Namibia. The few existing studies mainly focused on the analysis of prognostic variables of HIV and many other reports on HIV, which only provided the summary statistics of HIV prevalence at site or district level. Whereas almost non-existent studies on measles in Namibia are found in literature. In order to achieve an accurate estimation of measles and the HIV burden, this study intended to develop models for statistical analysis of spatially misaligned measles data and HIV data obtained from different sources fraught with selection bias and possible non-sampling errors for more robust maps of diseases burden in Namibia.

1.4 Objectives

1.4.1 Main objective

The main aim of this study is to develop models for disease mapping for spatial misaligned measles data while adjusting for measurement errors and while using multiple sources such as the national HIV sentinel surveillance (NHSS) and the Namibia demographic and health survey (NDHS) HIV data. In other words, the dissertation focuses in spatial and spatio-temporal methods that can help deal with misalignment and measurement errors in data, and model jointly data from two different sources.

1.4.2 Specific objectives

The specific objectives are to:

- fit models for misaligned data with application to map the risk of measles at sub-region level (i.e. constituency level) using data obtained at the regional level in Namibia;

- fit models for misaligned data fraught with measurements errors with application to map the spatio-temporal risk of measles at the sub-regional level (i.e. constituency level) using data obtained at the regional level in Namibia for the period 2005-2014; and

- develop joint models for national HIV sentinel surveillance (NHSS) and Namibia demographic and health survey (NDHS) HIV data.

1.5 Significance of the study

The multi-step modelling approach developed in this study would enable the use of regional aggregated data to build models that are useful for constituency level inferences. Also, this study is significant to Namibia as the health care providers and health policy makers would be in position of using the diseases' maps to efficiently deliver much needed care services across the country through identification of groups of people and or areas in needs. Furthermore, the study used a bivariate modelling approach that helped in dealing with spatially misaligned data and enhanced the prediction of HIV prevalence by jointly modelling DHS data source with other HIV data such as NHSS data.

1.6 Delimitations

With measles data, the period of 2001−2004 was excluded from the study period as the available information for this period was inconsistent throughout the country. Although the administrative boundaries have changed over time, this study had used the 2011 administrative boundaries (old boundaries) because they match with variables obtained from 2011 Namibia population and housing census.

1.7 Dissertation outline

Chapter 1 presented the overview, the conceptual framework, the background to spatial modelling, the orientation of the study, the statement of the research problem; and research objectives. In chapter 2, we begin with a short review of the basic Bayesian modelling. Different prior distributions as well as estimation methods, which include the classes of Markov Chain Monte Carlo (MCMC) and Integrated nested Laplace approximation (INLA) methods, are discussed. Furthermore, this chapter provides a review of spatial and spatio-temporal modelling approaches that have been applied for disease mapping. In addition, some current issues in spatial and spatio-temporal modelling are reviewed. The goal of this chapter is then to provide the reader with some general understanding of spatio-temporal modelling and lay some ground to the concepts and statistical methods used in disease mapping that can help the development of the subsequent chapters.

Chapter 3 discusses the problem of misalignment in data and it focuses on estimating and mapping the risk of measles at sub-regional level using data obtained at regional level in Namibia. To deal with misalignment in measles data from Namibia for the period 2005-2014, we proposed a multi-step approach to correct the misalignment.

Chapter 4 is devoted to spatio-temporal models involving Bernardinelli and Knorr-Held spatio-temporal models commonly used in space-time modelling. The multi-step approach was extended to spatio-temporal in order to account for temporal effects. Additionally, instead of the restrictive assumption that covariates remained constant over time, classical measurement error models in covariates were introduced to improve the spatio-temporal ecological regression model.

Chapter 5 provides the reviews of methods commonly used in multivariate disease mapping. Shared component models were adopted to fit bivariate models, using stochastic partial differential equations (SPDE) that help dealing with edge effects, for NDHS and NHSS surveys.

Chapter 6 revisits the primary objectives in order to evaluate if they have been achieved and it presents conclusions and recommendations for improvements and future studies. After Chapter 6, an appendix that includes all R-programmes used in this dissertation is provided. A biography of all references is given at the end of dissertation.

Chapter 2

Approaches to spatio-temporal analysis and disease mapping

2.1 Generalised linear models

2.1.1 Introdution

Linear regression model requires that the response variable must be continuous and normally distributed (Scott, 2007). However, in some research fields such as social science, continuous outcome variables are rare (Lindsey, 2001). Quite often, dichotomous, ordinal, or nominal outcomes are available (Gill, 2001). In these cases, the linear regression model becomes inappropriate due to several reasons that include heteroscedasticity and non-normal errors encountered in outcomes that are not continuous (Scott, 2007). The generalised linear modeling is a framework which provides a way to handle these problems. It provides a collection of models which relax the normality assumption of error terms to accommodate a wide range of error term distributions. Generalised linear models consist of three main components, namely: random, systematic, and link function (Waller & Gotway, 2004).

The random component consists of independent outcomes, denoted by Y_i for $i = 1, \ldots, n$, from a distribution within the exponential family. This implies that the probability density or mass function of Y_i may be expressed in the form

$$f(y_i, \phi_i) = exp\left(a(\phi_i) + b(y_i) + c(y_i)Q(\phi_i)\right), \tag{2.1}$$

where $a(\cdot)$, $b(\cdot)$, $c(\cdot)$, and $Q(\cdot)$ represent arbitrary functions of distributional parameters (ϕ_i) and observed values (y_i). If $c(y_i) = y_i$, the distribution is said to be in canonical form (i.e. standard form) (Dobson, 2002). In case of a vector of y_is independent values from the exponential family, then Eq. (2.1) can be generalised as follows (in canonical form).

$$f(y_1, \ldots, y_n, \phi_1, \ldots, \phi_1) = exp\left(\sum_{i=1}^{n} a(\phi_i) + \sum_{i=1}^{n} b(y_i) + \sum_{i=1}^{n} y_i Q(\phi_i)\right) \tag{2.2}$$

The systematic component of a generalised linear model is given by $X\beta$, where X is the design matrix, each row listing the values of covariates observed corresponding to the observation of the outcome y_i, and β is the vector of parameters.

The link function $h(\cdot)$ provides a functional connection between the systematic component $X\beta$ and the expected value of $\mathbf{Y} = (Y_1, \ldots, Y_n)$. That is

$$h[E(Y)] = X\beta \tag{2.3}$$

For members of the exponential family, the mean, $E(Y)$, is often among the distributional parameters ϕ and hence $Q(\phi_i)$ is often a function of $E(y_i)$ (Waller & Gotway, 2004). Table 2.1, adapted from (McCullagh & Nelder, 1989; Scott, 2007), provides some examples of standard models and their corresponding link functions.

2.1.2 Logistic and Poisson regression

Logistic and Poisson regression models have a number of applications in many fields such as epidemiology, social science, and disease mapping (Samaniego, 2010). These two families of generalized linear models are core to the work covered in this dissertation. Therefore, a review of these models in the subsequent sections is presented. Their parameters' estimation and model extension to spatial model follow the the frequentist and Bayesian spatial modeling framework as pointed out in the literature.

Table 2.1: Some examples of generalized linear models and corresponding link functions

Model	**link function** $h(\mu_i)$
Linear (Normal)	μ_i (identity)
Logistic	$log[\frac{\mu_i}{1-\mu_i}]$ (logit)
Binomial	$log[\frac{\mu_i}{1-\mu_i}]$ (logit)
Poisson	$log(\mu_i)$ (log)
Probit	$\Phi^{-1}(\mu)$ (inverse of the cumulative stadard normal)
log-log	$-log(-log(\mu_i)$ (log-log)
Complementary log-log	$log(-log(1-\mu_i))$ (c-log log)
Gamma	$\frac{1}{\mu_i}$
Inverse Gaussian	$\frac{1}{\mu_i^2}$ (Quadratic inverse:reciprocal2)
Negative binomial	$log(1-\mu)$ (log)

Logistic regression

Suppose that y_i is a binary outcome, where $y_i = 1$ indicates the presence of the disease of interest in individual i and $y_i = o$ denotes its absence. Let μ denote the unknown probability of disease prevalence in the population under study. The random variable Y_i follows a Bernoulli distribution with probability of disease μ. The likelihood function associated with the observations $y_1, \ldots, y_n$ is given by

$$f\left(y_1, \ldots, y_n, \mu\right) = \Pi_i^n \mu^{y_i} \left(1-\mu\right)^{1-y_i}. \tag{2.4}$$

The Eq. (2.4) can be rewritten as

$$f\left(y_1, \ldots, y_n, \mu\right) = exp\left[\sum_{i=1}^{n} log\left(1-\mu\right) + \sum_{i=1}^{n} y_i log\left(\frac{\mu}{1-\mu}\right)\right] \tag{2.5}$$

It can be noted that Eq. (2.5) is a member of an exponential family in canonical form with $\phi_i = \mu$, $a(\phi_i) = log(1-\mu)$,$b(y_i) = 0$, $c(y_i) = y_i)$, and $Q(\phi_i) = log\left(\frac{\mu}{1-\mu}\right)$. Since $E(Y_i = \mu)$, the canonical link is $h(E(Y_i) = h(\mu) = log\left(\frac{\mu}{1-\mu}\right)$ which is a logit link.

Poisson regression

The Poisson regression is commonly used as an approximation to the binomial distribution in modeling count data of rare diseases (Roussas, 1997; Dobson, 2002). It is also often used to model observed point locations as random events. Suppose that we observe location counts $y_1, \ldots, y_I$ that are independently and identically distributed Poisson random variables with mean and variance equal to $Var(Y_i) = E(Y_i = \mu)$. The joint probability associated with the observed data $y_1, \ldots, y_I$ is

$$f\left(y_1, \ldots, y_I, \mu\right) = \Pi_i^I \frac{\mu^{y_i} e^{-\mu}}{y_i!}, \tag{2.6}$$

which can be rewritten as

$$f(y_1, \ldots, y_I, \mu) = exp\left(\sum_{i=1}^{I} y_i log(\mu) - I\mu - \sum_{i=1}^{I} log(y_i!)\right) \tag{2.7}$$

The Poisson distribution is a member of an exponential family in canonical form with $\phi_i = \mu$, $a(\phi_i) = I\mu$,$b(y_i) = 0$, $c(y_i) = y_i$, and $Q(\phi_i) = log(\mu)$. Since $E(Y_i = \mu)$, the canonical link is $h(E(Y_i) = h(\mu) = log(\mu)$, which is a log link.

The estimation of parameters of generalised linear models is discussed in subsequent sections (i.e. frequentist and Bayesian approaches to estimation). It is worthwhile to mention that more attention was given to the Bayesian modelling approach as it forms the core of this research.

2.2 Frequentist modelling approach

2.2.1 Introduction

Under the frequentist modelling approach, data are often assumed to be a random sample of independently identically distributed (i.i.d) variables, it is argued that the i.i.d. assumption is logically shaky even though making this assumption may be relatively harmless (DeGroot, 1970). With this approach, model fits are often judged on the basis of their theoretical average performance. For example, estimators are compared on the basis of their mean squared root errors (MSEs). This means the comparison is based essentially on the on the squared error in many identical trials of an experiment. These averages might be inappropriate for assessing the merits of statistical procedures as identical repetitions of an experiment might be impossible (Samaniego, 2010). In addition to this, comparing the fit of non-nested models, such as a nonlinear model and its linearised version, may result in problems (Bolstad, 2004).

Generally, two methods are commonly used when dealing with estimation of parameters, namely: (i) optimizing relative to a risk-based criterion for a fixed sample size n and (ii) optimizing relative to some asymptotic measure of performance (as $n \rightarrow \infty$). Data with complex structures such hierarchical nesting of subjects, crossed classifications, spatially indexed data, or repeated measures on subjects present challenges (Congdon, 2010).

When sufficient data are available, the generalized least squares (GLS) method may be employed as it relaxes the assumptions of independent and constant errors by allowing autocorrelation and heteroscedasticity in residuals (Davidian & Giltinan, 1995).

In case of non-linear data, the GLS method becomes inapplicable. Two classes of iterative approximation methods, which involve the linearisation of non-linear models, are recommended. The first method is a first order linear approximation and it draws inferences from the joint maximum likelihood and generalised least squares methods. The second is a conditional first-order linear approximation seen as a refined version of the first order linear approximation (Davidian & Giltinan, 1995). The convergence of these methods depends on a number of factors that include starting values, overparameterisation, and the presence of parameters that may prevent the model to have a behavior of a close linear form (Davidian & Giltinan, 1995; Ratkowsky, 1990).

A good estimation method should ensure that an estimator satisfies, among others, the following properties: unbiasednes, effeciency, sufficiency, completeness, minimum variance, and best linear, and robustness (Wackerly et al., 2002). This section briefly reviews unbiasednes, sufficiency, completeness, and minimum variance. Details on the other properties can be found elsewhere (e.g. Samaniego (2010)).

2.2.2 Unbiasedness, sufficiency, and completeness of an estimator

Before the search of best estimators begins, the condition of unbiasedness for estimators should be met.
An estimator $\hat{\theta}$ is said to be unbiased of a parameter θ if

$$E(\hat{\theta}) = \theta \tag{2.8}$$

A parameter θ may have more than one competing unbiased estimators. The selection of an optimal estimator is achieved through the examination of sufficiency and completeness properties.
A statistic $U = g(y_1, \ldots, y_n)$ is a sufficient estimator of θ if the conditional distribution of the data $y_1, \ldots, y_n$, given U, does not depend on θ. Stated differently, a statistic $U = g(y_1, \ldots, y_n)$ is sufficient for estimation of unknown θ if and if the likelihood $L\left(y_1, \ldots, y_n \mid \theta\right)$ can be factorised in two non-negative functions as follows (Wackerly et al., 2002).

$$L\left(y_1, \ldots, y_n \mid \theta\right) = k\left(U, \theta\right) \times h\left(y_1, \ldots, y_n\right), \tag{2.9}$$

where $k(\cdot)$ is a function of U and θ, and $h\left(\cdot\right)$ is a function of the data only.

A sufficient statistic $U = g(y_1, \ldots, y_n)$ is said to be complete for the parameter θ if the equation $E_\theta(t(U)) = 0$ holds for a given function t, then $t(U) = 0$ with probability one (Roussas, 1997; Samaniego, 2010). This means that the completeness property ensures that there is only one function of the sufficient statistic U that is unbiased for θ.

2.2.3 Minimum variance unbiased estimators (MVUEs)

The estimator with the smaller variance is preferred as its average squared distance from the target parameter θ would be smaller than that of the other estimators. Generally, the estimation procedure would seek an estimator with the smallest possible variance. Two theoretical results that are commonly used in search of a minimum variance unbiased estimator, namely: Crao-Blackwell and Cramér-Rao theorems, are presented below (Samaniego, 2010).

Crao-Blackwell theorem:

Suppose that $Y_1, \ldots, Y_n$ are identically and independently distributed random variables, and that U, a function of $Y_1, \ldots, Y_n$, is a sufficient statistic for θ. Let $S = S(Y_1, \ldots, Y_n)$ be unbiased estimator of θ, and define $\hat{\theta} = \hat{\theta}(U) = E_\theta(S \mid U)$. Then the estimator $\hat{\theta}$ is unbiased for θ, and

$$Var_\theta(\hat{\theta}) \leq Var_\theta(S), \quad \forall \theta \in \Theta \tag{2.10}$$

This indicates that, while searching for good unbiased estimators in a given problem, one does not have to look beyond those that are functions of the sufficient statistic U, as any other unbiased estimator may be replaced by an unbiased estimator which is a function of U.

Cramér-Rao inequality:

Cramér-Rao theorem is expressed as an inequality and it provides a lower bound on the variance of unbiased estimators for a given problem. It holds under a set of conditions generally referred to as regularity conditions. Firstly, the support of the model is an open interval of real numbers which is independent of θ. Secondly, one may pass derivatives under integral signs as needed, provided the expectations exist. This is expressed mathematically as

$$\frac{\partial}{\partial \theta} \int s(x) f(x,\theta) dx = \int s(x) \frac{\partial}{\partial \theta} f(x,\theta) dx, \tag{2.11}$$

where $s(x)$ is any integrable function. The Cramér-Rao theorem relates the variance of the estimator to Fisher Information (Wackerly et al., 2002) . The Fisher $I_X(\theta)$, which reflects the information content about θ in a single observation x, is defined as

$$I_X(\theta) = E\left[\left(\frac{\partial}{\partial \theta} ln f(X,\theta)\right)^2\right] \tag{2.12}$$

Now the Cramér-Rao theorem is as follows

Suppose that $X_1, \ldots, X_n$ are independently and identically distributed, a probability distribution with density or probability mass function $f(x,\theta)$ for all θ in some open interval of real numbers.

Furthermore, let $\hat{\theta}$ be an unbiased estimator of θ based on $X_1, \dots, X_n$. Then

$$Var(\hat{\theta}) \geq \frac{1}{nI_X(\theta)} \tag{2.13}$$

2.3 Spatial modelling

2.3.1 Introduction

Spatial data are any form of data attached to geographical locations. In statistical literature, there are three main forms of spatial data, namely, areal data, point-referenced data, and point pattern data. Modelling such data can be considered from a hierarchical perspective, where the random effects are introduced to account for spatial dependence unexplained by the observed data. Modelling in the Bayesian framework is commonly preferred over the non-Bayesian approach that assumes parameters to be fixed but unknown, whereas the former considers all parameters to be stochastic. Hence, each parameter is assigned with a probability function known as a prior function. Likelihood models and prior functions are the most important parts of Bayesian inference. The likelihood function describes the dependence of a set of parameters on sample values and it is believed to portray in its totality the information contained in a data set, while the latter provides extra information about parameters through beliefs or assumptions as they are assigned before seeing data (Lawson, 2013). The conventional likelihood model formulation assumes that data are conditionally independent. This assumption allows formulating the likelihood function as a product of individual contributions of each observation y_i as follows. Let y_i, for i=$1, \dots, n$, be a sample of observed values. Then the likelihood of y_i is defined by

$$L(y_i|\theta) = \prod_{i=1}^{n} f(y_i|\theta), \tag{2.14}$$

where θ is a vector of parameters and $f(\cdot \mid \cdot)$ is a probability density function or a probability mass function.

Clearly, in a spatial context where spatial units are expected to obey Tobler's law of geography, the assumption of the conditional independence of data is violated. Locations that are nearby each other have very similar values relative to those located far from one another. This implies that, within spatial analysis, spatial correlation is crucial and must be catered for during analysis. It is accounted for in the prior probability distribution level, but not in the likelihood function.

2.3.2 Prior distributions

An important phase in Bayesian modelling is to choose prior distribution for parameters (Scott, 2007). When there is little information available, it is desirable to choose a prior distribution that does not dominate the likelihood function and such prior distributions are assumed not to have a strong influence and they are known as noninformative, or vague or reference or flat prior distributions (Lawson, 2013). Prior distributions based on the Fisher information matrix, known as Jeffrey's priors were developed in order to meet the criterion of noninformativeness. The literature also suggests that the choice of flat prior distributions can be guided by the understanding of the general behaviour of the variable of interest on its range. For instance, gamma, inverse gamma or uniform are quite often used as noninformative priors for variance parameters as such parameters are expected to be on a positive side of the real number line. Whilst, for parameters expected to be on either side of zero of a real number line, prior distribution with mean zero and large variance, Laplace distribution are natural choices. Other important criteria considered when choosing priors are impropriety and conjugacy (Bolstad, 2004; Scott, 2007). The former criterion is defined as prior distribution, which when its integral is evaluated over its range is not finite. This may result in an improper posterior. The latter criterion can be defined as a combination of prior distributions and likelihood functions that yield posterior distributions which are in the same distribution family with their prior distributions. This property ensures that the integration of a posterior distribution over its range is finite and hence analytical methods can be used for the evaluation of such a posterior distribution. A brief discussion of some commonly used conjugate priors are provided below.

Beta and uniform priors

Let θ provided in Eq. (2.14) be a random variable representing a proportion. A convenient class of density functions for θ is a beta density function defined by

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \tag{2.15}$$

where α and β represent the number of prior successes and failures, respectively; $\Gamma[\cdot]$ is a gamma function.

The choice of the parameters α and β depends on the amount of available information and the magnitude of weight the available information should carry in the posterior inference (Scott, 2007). That is, α and β will assume large values if large prior information is available and it is judged to be very important relative to current data. In case there is no sound reason as to why much importance should be put into the prior information, then α and β are given moderate values. If very non-significant prior information is available or a researcher wishes to assign very little weight to prior information, then small values are given to the parameters. The commonly used small value is $\alpha = \beta = 1$. In this case, it can be easily shown that Eq. (2.15) simplifies to

$$p(\theta) = 1, 0 \leq \theta \leq 1, \tag{2.16}$$

which is a uniform prior distribution. Both beta and uniform distribution are prior conjugates of a binomial distribution.

Gamma and inverse gamma priors

For count data, a Poisson probability mass function is commonly used. In this case, the objective is to assign a prior distribution to rate parameter θ. A gamma distribution, which is a conjugate prior for Poisson distribution, is a natural choice for θ and it is expressed as follows

$$p(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \tag{2.17}$$

where α and β are shape and scale parameters, respectively.

An inverse gamma distribution for variable ϕ, provided that $\frac{1}{\phi}$ follows a gamma distribution, is given by

$$p(\phi) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\phi^{-(\alpha+1)}e^{-\frac{\beta}{\phi}}; \phi, \beta, \alpha > 0 \tag{2.18}$$

Gamma and inverse gamma are generally used as conjugate priors for the precision parameter ($\frac{1}{\sigma^2}$) and variance (σ^2), respectively, in normal distribution.

Dirichlet prior distribution

A Dirichlet distribution is a multivariate extension of beta distribution. If θ is a vector of r dimension and θ follows *Dirichlet* (α_1,...,α_r), then

$$p(\theta) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_r)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_r)}\theta_1^{\alpha_1-1}\ldots\theta_r^{\alpha_r-1}, \tag{2.19}$$

where α_1,...,α_r represent prior totals of outcomes in each of the r outcome categories. Since the multinomial distribution is a generalisation of a binomial distribution and a beta distribution is a conjugate prior for a binomial distribution, consequently the Dirichlet distribution is a conjugate prior for a multinomial distribution.

Wishart and inverse wishart priors

If parameter $\phi \sim$ Wishart (S), then

$$p(\phi) \propto |\phi|^{\frac{(\nu-d-1)}{2}} \exp(-\frac{1}{2}tr(S^{-1}\phi)) \tag{2.20}$$

Also, if parameter $\phi \sim$ inverseWishart (S), then

$$p(\phi) \propto |\phi|^{-\frac{(\nu+d+1)}{2}} \exp(-\frac{1}{2}tr(S\phi^{-1})) \tag{2.21}$$

S is a d dimension scale matrix and ν is the number of degrees of freedom; S and ϕ are assumed to be positive definite (i.e. $z^T S z > 0$ and $z^T \phi z > 0$, for a non-zero z vector of dimension d). Wishart and inverse Wishart are generalisations of gamma and inverse gamma in multivariate normal distribution. Consequently, the inverse Wishart is a conjugate prior for variance-covariance matrix in multivariate distribution.

Normal prior

If $X \sim Normal(\mu,\ \sigma^2)$, then μ is usually assigned a normal flat prior (i.e $\mu \sim N(0,\ \sigma_\mu^2)$) and $\sigma^2 \sim IG(\alpha,\ \beta)$. In case of multivariate normal distribution, a multivariate normal prior and inverse Wishart prior are used for the vector of means μ and the variance covariance matrix, respectively.

2.3.3 Posterior distribution

Recall that in Bayesian analysis, both parameters θ and data y are random variables. Let $f(\theta,\ y)$ be the joint distribution of θ and y. Then $f(\theta,\ y)$ can be rewritten as $f(\theta,\ y)= f(y \mid \theta)p(\theta)$, where $p(\theta)$ is the prior distribution of θ and $f(y \mid \theta)=L(\theta)$ is the likelihood function. Also, it can be shown that $f(\theta,\ y)=p(\theta \mid y\ )f(y)$,where $p(\theta \mid y\ )$ is the conditional distribution of θ given the observed data y and $f(y)$ is a marginal distribution y. Thus, $f(y)$ can be derived from the joint as follows $f(y)=\int f(y \mid \theta)p(\theta)d\theta$.

Now by making $p(\theta \mid y\ )$ the subject of the formula from the joint distribution, it follows that

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int f(y|\theta)p(\theta)d\theta} = \frac{L(\theta)p(\theta)}{\int L(\theta)p(\theta)d\theta} \tag{2.22}$$

It can be noticed that the posterior distribution derived in Eq. (2.22) is proportional to the product of the likelihood and the prior distribution where $\int f(y \mid \theta)p(\theta)d\theta$ is the constant of proportionality. Therefore, the posterior distribution of θ can be expressed differently as

$$p(\theta|y) \propto L(\theta)p(\theta) \tag{2.23}$$

Table 2.2 provides a summary of priors, likelihoods and their conjugate posteriors.

Table 2.2: Some of the commonly used priors and their conjugate posteriors (Congdon, 2010; Ntzoufras, 2009)

Prior distribution: $\pi(\cdot)$	Likelihood: $f(Y \mid \cdot)$	Posterior distribution:$\pi(\cdot \mid Y)$
Gam:$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$	Poi:$f(Y \mid \theta) = \frac{e^{-n\theta}\theta^{\sum y_i}}{\prod_{i=1}^{n} y_i!}$	$\pi(\theta \mid Y) \propto e^{-(n+b)\theta}\theta^{n\bar{y}+a}$
Beta:$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$	Bin:$f(Y \mid \theta) = \prod_{i=1}^{n} N_iC_{y_i}\theta^{n\bar{y}}(1-\theta)^{N-n\bar{y}}$	$\pi(\theta \mid Y) \propto \theta^{n\bar{y}+\alpha-1}(1-\theta)^{N-n\bar{y}+\beta-1}$
Dir:$\pi(\theta) = \frac{\Gamma(\alpha_1+\ldots+\alpha_r)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_r)}\theta_1^{\alpha_1-1}\ldots\theta_r^{\alpha_r-1}$	Mul: $f(Y \mid \theta) = \frac{y!}{\prod_{i=1}^{r} y_{i=1}!}\prod_{i=1}^{r} beta_i^{y_i}$	$\pi(\theta \mid Y) \propto \prod_{i=1}^{r} \theta_i^{\alpha_i+y_i-1}$
N-G: $\pi(\mu, \sigma^{-2}) = N(\mu_0, c\sigma^{-2})G(\alpha, \beta)$	N:$f(Y \mid \mu, \sigma^2 = (2\pi\sigma^2)^{\frac{-n}{2}}exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\}$	$\pi(\mu, \sigma^{-2} \mid Y) \propto \sigma^{-(n+1)+\alpha-1}\times$ $exp\{-\frac{\sigma^2}{2}(\frac{1+nc}{c})(\mu - \frac{\mu_0+nc\bar{y}}{1+nc})^2 - \sigma^{-2}(\beta + \frac{K}{2}\},$ $K = \sum_{i=1}^{n} y_i^2 + \frac{\mu_0^2}{c} - \frac{(nc\bar{y}+\mu_0)^2}{c(nc+1}$
N-G: $\pi(\mu, \sigma^{-2}) = N(0, \sigma_\mu^{-2})G(\alpha, \beta) \propto \sigma^{-2}$	N:$f(Y \mid \mu, \sigma^2 = (2\pi\sigma^2)^{\frac{-n}{2}}exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\}$	$\pi(\mu, \sigma^{-2} \mid Y) \propto \sigma^{-(n+2)}exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\}$

2.3.4 Estimation methods

After a posterior distribution has been obtained, the next step is to derive summary measures of interest from the posterior distribution. For relatively simple posterior distributions, two methods of sampling, namely inversion and rejection sampling methods, can be used to obtain samples. The first method is executed in two main steps. In the first step, a random variable u is drawn from $U(0,\ 1)$. Then, in the second step draw z=F^{-1} (u) from the posterior distribution. The second method is implemented in three basic steps: firstly, draw a value z from any easy envelop distribution of the posterior distribution; secondly, compute the ratio of the envelop distribution evaluated at the value obtained in step 1 to the posterior distribution evaluated at the same value; lastly, draw a random variable u from U(0,1) and accept z if the ratio is greater than u. In most cases, it is not straight forward to obtain summary measures through these methods or direct maximisation or analytical means of the posterior distribution (Scott, 2007; Lawson, 2013). The complexity of the posterior distribution of parameters that results from hierarchical levels of models requires the use of advanced sampling algorithms. These algorithms enable to derive samples from the posterior distributions, which in turn are summarized to produce the parameter estimates of interest. Two major classes, namely Markov Chain Monte Carlo (MCMC) and Integrated Nested Laplace Approximation methods are reviewed in the subsequent sections.

2.3.4.1 Markov Chain Monte Carlo Methods (MCMC)

Markov Chain Monte Carlo Methods is a collection of iterative simulation methods. These methods include Metropolis-Hastings, Gibbs sampler, hit-and-run sampler, shake-and-bake algorithm, Metropolis-Gibbs hybrids and the multiple-try Metropolis-Hastings method, auxiliary variable samplers (e.g. the slice sampler and the Swendsen-Wang algorithm), and reversible-jump sampler (Kroese et al., 2011). They involve the process of sampling a new value from the posterior distribution, given the availability of previous values through the random simulation process. For the past two decades, Markov Chain Monte Carlo (MCMC) methods that include the Gibbs sampler and the Metropolis-Hastings algorithm have become popular methods to sample from complex or intractable posterior distributions. Their popularity is due to the fact that they can sample from uni or multi dimensional posterior distributions and navigate through the whole support of the posterior distribution. However, the success of MCMC methods depends on the choice of a good proposal distribution, which allows reasonably quick movement across the support of the posterior. In subsequent paragraphs, we briefly review Gibbs sampler and Metropolis-Hastings algorithm. More details on these methods can be found elsewhere (e.g. Robert & Casella (2011); Kroese, Taimre, & Zadravko (2011); Scott (2007); Tierney (1994)).

Metropolis-Hastings (MH) algorithm

In MH algorithm, samples are produced from a probability distribution using the full joint density function. One distinct advantage of MH algorithm relative to other sampling methods is that it works with multivariate distributions and it does not require an envelop function. An MH algorithm consists of the iteration of the following four steps:

1. Choose starting values for the vector of parameters θ, say ϕ .

   A careful selection of starting values is encouraged as poor starting values may cause the algorithm not to move fast towards the main support of the posterior.

2. Choose a proposal density $\alpha[\cdot]$ from which a candidate value for the parameter θ^c will be simulated.

   Although asymmetric proposal densities still work, it is recommended to choose symmetric proposal densities.

3. Compute the ratio $R=\frac{p(\theta^c)\alpha(\theta^{j-1}|\theta^c)}{p(\theta^{j-1})\alpha(\theta^c|\theta^{j-1})}$.

   If an asymmetric proposal density is employed, some candidate values may be selected more often than others. To deal with this problem, a correction measure is introduced in the ratio expression (i.e. the ratio of the proposal densities evaluated at the candidate and previous points. That is $\frac{\alpha(\theta^{j-1}|\theta^c)}{\alpha(\theta^c|\theta^{j-1})}$).

4. Draw a random variable u from $U(0, 1)$ and compare it with the R to determine whether the candidate value is from the target distribution or not.

   If $R > u$, then the candidate is accepted as a draw from the posterior density

$p(\cdot)$. Otherwise, the previous parameter value is retained. That is if $R > u$, then set $\theta^j = \theta^c$. Otherwise, set $\theta^j = \theta^{j-1}$

Gibbs sampler

Gibbs sampler is a particular case of the Metropolis-Hastings algorithm. It is suitable when the MH algorithm fails to sample from a high dimensional posterior distribution. The Gibbs sampler algorithm breaks the complex posterior into a series of simple conditional distributions from which it is feasible to sample (Robert & Casella, 2010). The description of a basic multistage Gibbs sampler is given below. Suppose that, for some $k > 1$, θ is a vector of random variables which can be written as θ=(θ_1,. . .,θ_k). Further, assume that it is possible to simulate from the corresponding conditional densities $p_1,\ldots,p_k$.
In other words, $\theta_i \mid \theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_k \sim p_i(\theta_i \mid \theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_k)$, for i=1,. . .,k

1. Assign a vector of staring values, ϕ, to the parameter vector (i.e. ϕ=$\theta^{j=0}$)

2. Set j=j+1 , j is the iteration counter

3. Simulate $\theta_1^{(j)} \mid \theta_2^{j-1},\ldots,\theta_k^{j-1} \sim p_1(\theta_1^{(j)} \mid \theta_2^{j-1},\ldots,\theta_k^{j-1})$

4. Simulate $\theta_2^{(j)} \mid \theta_1^{j},\theta_3^{j-1},\ldots,\theta_k^{j-1} \sim p_2(\theta_2^{(j)} \mid \theta_1^{j},\theta_3^{j-1},\ldots,\theta_k^{j-1})$
   $\vdots$

5. Simulate $\theta_k^{(j)} \mid \theta_1^{j},\theta^{j},\ldots,\theta_{k-1}^{j} \sim p_2(\theta_k^{(j)} \mid \theta_1^{j},\theta^{j},\ldots,\theta_{k-1}^{j})$

6. Return to step 2 to start a new loop

An iteration of the Gibbs sampler is a loop through these steps and every loop gives rise to a new sampled value of a parameter called an updated value.

2.3.4.2 Integrated Nested Laplace Approximation (INLA)

Markov Chain Monte Carlo (MCMC) methods described in the section above are simulation-based methods used for Bayesian computation. Although MCMC methods are tremendously flexible and capable to deal with almost any type of data and model, these methods are computationally expensive to obtain the posterior distribution for the parameters (Blangiardo, Cameletti, Baio, & Rue, 2013). Consequently, Integrated Nested Laplace Approximation, which is an analytic approximation based on the Laplace method, has been recently developed as an alternative to MCMC. In the subsequent section, we provide a description of the Laplace approximation method and describe the INLA algorithm. In depth details can be obtained in Blangiardo et al. (2013); Blangiardo & Cameletti (2015).

Laplace analytic approximation method

Let $p(\theta)$ be a posterior distribution of a random variable of parameters θ. The main purpose of a Bayesian inference is to evaluate the integral:

$$\int p(\theta)d\theta = \int exp(logp(\theta))d\theta \tag{2.24}$$

Applying the Taylor expansion to log p(θ) and evaluating the expansion at θ =θ_0, it follows that

$$logp(\theta) \approx log(p(\theta_0)) + (\theta - \theta_0)\frac{\partial log(p(\theta))}{\partial \theta} \mid_{\theta=\theta_0} + \frac{(\theta-\theta_0)^2}{2}\frac{\partial^2 log(p(\theta))}{\partial \theta^2} \mid_{\theta=\theta_0} \tag{2.25}$$

It can be shown that $\frac{\partial log(p(\theta))}{\partial \theta} \mid_{\theta=\theta^*}= 0$,where θ^*=mode.
Thus, Eq. (2.25) simplifies to

$$log(p(\theta)) \approx log(p(\theta^*)) + \frac{(\theta-\theta^*)^2}{2}\frac{\partial^2 log(p(\theta))}{\partial \theta^2} \mid_{\theta=\theta^*} \tag{2.26}$$

Therefore, the integral of the posterior distribution (Eq. (2.24)) can be approximated as follows.

$$\int p(\theta)d\theta \approx \int exp(log(p(\theta^*)) + \frac{(\theta-\theta^*)^2}{2}\frac{\partial^2 log((\theta))}{\partial \theta^2} \mid_{\theta=\theta^*})d\theta \tag{2.27}$$

With little algebraic manipulation, it can be shown that

$$\int p(\theta)d\theta \approx p(\theta^*) \times \int exp(\frac{(\theta-\theta^*)^2}{2\sigma^{*2}}), where\sigma^{*2} = -\frac{1}{\frac{\partial^2 log(p(\theta))}{\partial \theta^2} \mid_{\theta=\theta^*}} \tag{2.28}$$

It can be noted that $exp(\frac{(\theta-\theta^*)^2}{2\sigma^{*2}})$ is associated with the desnsity of a normal distribution with mean θ^* and variance -$\frac{1}{\frac{\partial^2 log(p(\theta))}{\partial\theta^2}|_{\theta=\theta^*}}$.
Therefore, the integral can be approximated using the cumulative distribution of the normal distribution. Say, for example if one wishes to evaluate the posterior distribution between two limits a and b, it can be achieved as follows

$$\int_a^b p(\theta)d\theta \approx p(\theta^*)\sqrt{2\pi\sigma^{*2}}(\Phi(b) - \Phi(a)) \tag{2.29}$$

INLA approximations of parameters and hyperparameters

Let $\mathbf{y}$=(y_1,...,y_n) be a vector of observed values. Generally, the distribution of y_i is defined by the additive linear predictor η_i, which is defined as

$$\eta_i = h(\phi_i) = \alpha_0 + \sum_{k=1}^{K} \alpha_k x_{ki} + \sum_{r=1}^{R} f_r(z_{ri}), \tag{2.30}$$

where $h(\cdot)$ is an appropriate link function, ϕ_i=$E(y_i)$, α=(α_0, α_1,..., α_K) is a vector of coefficients associated with the linear covariates $\mathbf{x}$=$(1, x_1, \ldots, x_K)$, and $\mathbf{f}$=(f_1, ...,f_R) is a vector of functions associated with the vector of covariates $\mathbf{z}$=(z_1, ...,z_R). The function $f_i(\cdot)$ may take various forms that include nonlinear effects, random intercept and slopes, temporal and spatial random effects. Let θ=(α, f) and Ψ=($\Psi_1, \ldots, \Psi_m$) be a vector of parameters and a vector of hyperparameters, respectively. Then the likelihood of n-observed data points given θ and Ψ is given by

$$p(y \mid \theta, \Psi) = \prod_{i}^{n} p(y_i \mid \theta_i, \Psi) \tag{2.31}$$

Assuming a multivariate normal on θ with a mean zero and a sparse precision matrix $Q(\Psi)$, then the posterior distribution of θ given Ψ is expressed as

$$p(\theta \mid \Psi) = (2\pi)^{-\frac{n}{2}} \mid Q(\Psi) \mid^{\frac{1}{2}} \exp(-\frac{1}{2}\theta^t Q(\Psi)\theta) \tag{2.32}$$

The joint posterior distribution of parameters and hyperparameters given the observed data is given by

$$p(\theta, \Psi \mid y) \propto p(\Psi) \mid Q(\Psi) \mid^{\frac{1}{2}} exp(-\frac{1}{2}\theta^t Q(\Psi)\theta + \sum log(p(y_i \mid \theta_i, \Psi)) \tag{2.33}$$

With Bayesian inference, the major objectives are the marginal posterior distribution of elements of the parameter vector θ and elements of hyperparameter vector Ψ. Mathematically, the objective is to evaluate

$$p(\theta_i \mid y) = \int p(\theta, \Psi \mid y) d\Psi = \int p(\theta_i \mid \Psi, y) p(\Psi \mid y) d\Psi \tag{2.34}$$

From the Eq. (2.34), it can be noted that $p(\Psi \mid y)$ and $p(\theta_i \mid \Psi, y)$ are unknown and hence have to be computed. The Laplace approximation is used to compute these posterior distributions of the hyperparaters and parameters, respectively as follows. First, computation of an approximation of $p(\Psi \mid y)$:
Using Eq. (2.34), we obtain that

$$p(\Psi \mid y) = \frac{p(\theta, \Psi \mid y)}{p(\theta \mid \Psi, y)} \tag{2.35}$$

Applying the law of conditional probability and law joint probability on the numerator of the left hand side of Eq. (2.35), we obtain

$$p(\Psi \mid y) = \frac{p(y \mid \theta, \Psi) p(\theta \mid \Psi) p(\Psi)}{p(y)} \frac{1}{p(\theta \mid \Psi, y)} \propto \frac{p(y \mid \theta, \Psi) p(\theta \mid \Psi) p(\Psi)}{p(\theta \mid \Psi, y)} \tag{2.36}$$

Substituting $p(\theta \mid \Psi, y)$ by its Laplace approximation, the Eq. (2.36) becomes

$$p(\Psi \mid y) \approx \frac{p(y \mid \theta, \Psi) p(\theta \mid \Psi) p(\Psi)}{\tilde{p}(\theta \mid \Psi, y)} \mid_{\theta=\theta^*(\Psi)} =: \tilde{p}(\Psi \mid y), \tag{2.37}$$

where $\theta^*(\Psi)$ is the mode for some given value of Ψ.

Second, computation of a Laplace approximation to the posterior distribution of each parameter ($p(\theta_i \mid \Psi, y)$): Similar mathematical manipulations are also applied to compute an approximation to the posterior distribution of each parameter and the resulting approximation is

$$p(\theta_i \mid \Psi, y) \approx \frac{p(\theta, \Psi \mid y)}{\tilde{p}(\theta_{-i} \mid \theta_i, \Psi, y)} \mid_{\theta_{-i}=\theta^*_{-i}(\theta_i, \Psi)} =: \tilde{p}(\theta_i \mid \Psi, y), \tag{2.38}$$

where $\tilde{p}(\theta_{-i} \mid \theta_i, \Psi, y)$ is the approximation of $p(\theta_{-i} \mid \theta_i, \Psi, y)$ through the Laplace approach, and $\theta^*_{-i}(\theta_i, \Psi)$ is the mode.
Consequently, by substituting Eq. (2.37) and Eq. (2.38) into Eq. (2.34), the marginal posterior distribution is approximated by

$$\tilde{p}(\theta_i \mid y) \approx \int \tilde{p}(\theta_i \mid \Psi, y) \tilde{p}(\Psi \mid y) d\Psi. \tag{2.39}$$

Numerical methods (e.g. finite weighted sum) are used to evaluate the Eq. (2.38). The interested reader may find more details on weighted sum methods elsewhere (e.g. Blangiardo et al. (2013)).

2.3.5 Spatial modelling of areal or lattice data

2.3.5.1 Areal-lattice data and spatial proximity measures

Areal-lattice data

Data whose locations in space are known and are realisations of a stochastic process indexed by space are called spatial data. That is $Y(s_i) \equiv \{y(s_i), s_i \in D\}$, where s_i is an areal unit with well defined boundaries in a fixed d-demensional space D, and $y(s_i)$ is random aggregate value over s_i. If s_i is irregular (usually based on administrative boundaries), then $Y(s_i)$ are called area data. Otherwise, $Y(s_i)$ are lattice data. For simplicity in our notations, we replace the vector of areal units $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ by the vector of indexes of areal units $(1, 2, \ldots, n)$.

Spatial proximity measures

A spatial proximity measure quantifies the spatial dependence between areas i and j in form of weights. It specifies the neighbourhood structure over an entire study domain. When the weights (b_{ij}) are collected in matrix form, it originates a spatial proximity matrix also known as a spatial connectivity matrix with the binary spatial connectivity matrix being the commonly used. Below are some functions defining the proximity weights (elements) of a binary spatial connectivity matrix (for more details see for example Waller & Gotway (2004)).

$$b_{ij} = \begin{cases} 1 & \text{if areas } i \text{ and } j \text{ share a boundary} \\ 0 & \text{otherwise} \end{cases} \tag{2.40}$$

where $b_{ii} = 0$. Eq. (2.40) specifies the neighbourhood in terms of areas that are adjacent. Other types of neighbourhoods based on idea of closeness, not necessarily on adjacency, can be defined as follows

$$b_{ij} = \begin{cases} 1 & \text{if the centroid of area } j \text{ is one of the } q \text{ nearest to the centroid of area } i \\ 0 & \text{otherwise} \end{cases} \tag{2.41}$$

In this case, the resulting spatial weights matrix is not necessarily symmetric as b_{ij} may not necessarily equal to b_{ji}. The idea of q nearest neighbours can be modified by defining the neighbours in relation with some parametric function of distance. This concept is based on a pre-defining disk smoothing window centered at the centroid of each area with an arbitrary radius δ.

$$b_{ij} = \begin{cases} 1 & \text{if } d_{ij} < \delta \\ 0 & \text{otherwise} \end{cases} \tag{2.42}$$

where d_{ij} is the euclidean distance between the centroids of areas i and j. An inverse power function of the distance d_{ij} can be used to yield the following elements of the spatial proximity matrix.

$$b_{ij} = \begin{cases} d_{ij}^{-\alpha} & \text{if } \alpha > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.43}$$

A neighbourhood structure based on the length of the boundary shared by areas i and j can be defined as

$$b_{ij} = \begin{cases} \frac{p_{ij}}{p_i} & \text{if areas } i \text{ and } j \text{ share a boundary} \\ 0 & \text{otherwise} \end{cases} \tag{2.44}$$

where p_{ij} is the length of the boundary common to areas i and j, and p_i is the perimeter of area i.

Measures of spatial autocorrelation

In general, it is practical to assume that areas that are close in space show more similarity than areas that are far apart. The measure of the strength of spatial similarity between areas can be quantified using spatial autocorrelation. There are many measures of spatial autocorrelation.

- Moran's I:

The Moran's I statistic is algebraically expressed as

$$I = \frac{1}{S^2} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}}, \tag{2.45}$$

where the expected value of Moran's I is given by $E(I)=-\frac{1}{n-1}$ and S^2 is the sample variance of observed y_i.

Moran's I statistic is the most popular choice to measure the spatial autocorrelation (Huo et al., 2011; Kamdem et al., 2012; Paireau, Girond, Collard, Maïnassara, & Jusot, 2012). It is a valid measure of spatial dependence under the null hypothesis that the related neighbours co-vary in no consistent way (i.e. randomness). It is generally used to measure spatial autocorrelation of continuous data though it can be used to analyse count data (Pfeiffer et al., 2008). Unlike the Pearson correlation coefficient, Moran's I statistic needs not to lie between -1 and 1 (Waller & Gotway, 2004; Banerjee, Carlin, & Gelfand, 2004). Its theoretical expression for the upper bound can be found in Waller & Gotway (2004)). It can be used to determine the strength of spatial dependence, distinguish between positive and negative spatial autocorrelation, detect spatial clusters and spatial outliers, and test the significance of the spatial correlation. High positive Moran's I value indicates possible clusters, low negative Moran's I value indicates that high and low values are interspersed, and a zero value of Moran's I statistic indicates the non-existence of spatial autocorrelation. It is worthy to note that, according to Waller & Gotway (2004), the spatial structure in the population sizes can induce measurable positive spatial correlation among the observed counts even if the constant risk hypothesis is met. For the Moran's I to reflect a true spatial pattern instead of a heterogeneous population distribution, it is advisable to use areal incidence rates, instead of areal disease counts (Waller & Gotway, 2004; Pfeiffer et al., 2008). Two types of this statistic are distinguished: Global Moran's I statistic and local Moran's I statistic. The former statistic is a useful measure of the overall clustering; whereas the latter statistic is an important tool for detecting local spatial patterns.

- Geary's C:

The Geary's C statistic is agebraically given by the form

$$C = \frac{1}{S^2} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}(y_i - y_j)^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}} \tag{2.46}$$

The Geary's C is a spatial analogue of Durbin-Watson statistic and the variogram for measuring the association in time series and in area of geostatistics, respectively (Banerjee et al., 2004). It measures the similarity between pairs of areal units but not similarity between neighbouring areas. C is always positive ($0 \leq C \leq 2$) and asymptotically normal if the y_is are identically and independently distributed (i.i.d). Small values (i.e. $0 \leq C \leq 1$) indicate positive spatial association whereas values greater than one suggest negative spatial dependence.

2.3.5.2 Gaussian Markov Random Fields (GMRF)

Let θ be a vector of parameters of interest in the space D. Generally, θ is a Gaussian random field if it is a Gaussian distributed random vector, which satisfies some conditional independence properties. That is, for any pair (i,j) such that $i \neq j$, then $\theta_i \perp \theta_j \mid \theta_{-\{i,j\}}$.

Let G=(ν, ε) represent a graph with ν =$\{1, \ldots, n\}$ a set of vertices,
and ε =$\{\{i, j\} : i, j \in \nu\}$ the set of edges in the graph; then for i, $j \in \nu$ the conditional independence holds if the edge $\{i, j\} \notin \varepsilon$, and does not hold otherwise.

Specifically, the definition of GMRF can be extended to reflect the condional independence properties which are in agreement with some specified G graph and a symmetric and positive definite (SPD) precision matrix Q as follows.

A random vector of parameters θ=$(\theta_i, \ldots, \theta_n)^T \in \mathbb{R}^n$ is called a GMRF with respect to labelled graph G=(ν,ε) with mean μ and SPD precion matrix Q, if its density has the form

$$f(\theta) = (2\pi)^{\frac{1}{2}}|Q|^{\frac{1}{2}} \exp(-\frac{1}{2}(\theta - \mu)^T Q(\theta - \mu)), \; and \quad Q_{i,j} \neq 0 \Longleftrightarrow \{i, j\} \in \varepsilon, \forall \quad i \neq j. \tag{2.47}$$

Further details on conditional independence properties and Markov properties of a GMRF can be found in (Gelfand, Diggle, Fuentes, & Guttorp, 2010).

The Bayesian framework that takes into account for spatial similarities based on neighbourhood structure is used to model areal data. For an area i, its first order neighbours $N(i)$ are the areas that share borders with it.
Thus, structure matrix R can be defined as

$$R = \begin{cases} N_i & \text{if } i = j \\ 1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

where $i \sim j$ indicates that area i is a neighbour of area j. Let θ=(θ_i,θ_{-i}) be a vector of parameters of interest in the space D such that θ_i of the ith area is conditionally independent of all other parameters given the set of its neighbors $N(i)$(i.e. $\theta_i \perp \theta_{-i} \mid \theta_{N(i)}$). Then a sparse precision matrix Q of θ, which is a function of the structure matrix R can be constructed in a way that for any pair of elements (i,j) in θ
$\theta_i \perp \theta_j \mid \theta_{-ij} \iff Q_{ij} = 0$. This is a specification of a Gaussian Markov random field (GMRF) (Blangiardo & Cameletti, 2015).

2.3.6 Spatial modelling of geostatistical data

Suppose that $y(s)$, with $s \in D \subset \mathbb{R}^2$, is a variable, which is in theory defined at every point over a bounded study region of interest (D) and has been observed at each of n distinct points. Most of the times, the data are fragmentary and often sparse. Therefore, the primary objectives of geostatistical modelling are to make inferences about the process that governs the spatial distribution of the variable and about values of the variable at unsampled locations. Geostatistical modellling has been applied in many different fields, such as mining, agriculture, fisheries, hydrology, geology, meteorology, petroleum, remote sensing, soil science and so on (Fischer & Getis, 2010).

2.3.6.1 Definitions

In this subsection, we begin by defining some of the key concepts that will be consistently used.

Point-referenced data and Gaussian field

Let $\{y(s), s \in D \subset \Re^2\}$ be a random field characterised by a spatial index s which varies continuously in the fixed study domain of two-dimensional space. Then the *point-referenced* data represents a data set resulting from the random field. Though these data naturally arise as realisations at a particular number of observation locations, theoretically they are assumed to be measured anywhere in study domain (Bivand, Pebesma, & Gomez-Rubio, 2008).

A random field $\{y(s), s \in D \subset \mathbb{R}^2\}$ is a *Gaussian field* if the vector of its realisations $y(s)$=$(y(s_1), \ldots, y(s_n))$ at n locations (i.e. s_1,. . .,s_n) is a multivariate normal random variable with mean $\mu = (\mu_1, \ldots, \mu_n)$ and spatial covariance matrix Σ. Σ_{ij}= $Cov(y(s_i), y(s_j))$=$C(y(s_i), y(s_j)$, where the covariance function $C(\cdot, \cdot)$ can assume different forms.

Stationary and isotropic processes

A stochastic process is *strictly stationary* when it is invariant to translation within d-dimensional space $\mathbb{R}^d$, usually $d = 2$. That is, for any collection of n areal units (s_1,. . .,s_n) and any separation (distance)vector $h \in \mathbb{R}^d$, the distribution of $y(s)$=$(y(s_1), \ldots, y(s_n))$ is the same as the distribution of $y(s{+}h)$=$(y(s_1{+}h), \ldots, y(s_n{+}h))$. A process becomes *weakly stationary or second order stationary* if it has a constant mean ($\mu(s) = \mu$) and the covariance is a function of only the difference between locations s and $s + h$ but not on the locations themselves (i.e. Cov($y(s)$,$y(s + h)$)=$C(h)$ for all h and s, $s + h \in D$).

An *isotropic process* is a process which is invariant to rotation about the origin. In other words, the relationship between any two events is only a function of the distance between the two events but not a function of the direction.

Otherwise, if a random process is a function of both distance and direction, it is called *anisotropic*.

Intrinsically stationary process, variogram and semivariogram

Let $\{y(s), s \in D \subset \mathbb{R}^2\}$ be a random field. Then $y(\cdot)$ is said to be *intrinsically stationary* if $E(y(s))= \mu \ \forall \ s \in D$(i.e. the process has a constant mean)and $Var\{y(s_i) - y(s_j)\}$=$2\gamma(h)$ (i.e. variance depends only on the euclidean distance between locations s_i and s_j but not on the locations themselves). In the literature, the functions $2\gamma(\cdot)$ and $\gamma(\cdot)$ are referred to as *variogram* and *semivariogram*, respectively. However, some scholars still refer to the plots of $2\gamma(\cdot)$ and $\gamma(\cdot)$ versus the spatial lag (h) as variogram and semivariogram, respectively. It can be easily shown that the semivariogram is a function of covariance at spatial lag h (i.e. $\gamma(h)$= $C(0) - C(h)$)).

2.3.6.2 Estimation of covariance and semivariogram

Assume that y(s) is a random variable observed at a number of spatial locations and whose characteristics are to be modelled within a given domain of study D (i.e. regionalised random variable).

Covariance

Generally, the covariance between any two variables, say x and y, measures the degree to which x co-varies with y. It is calculated as
$C(x, y)$=$\frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$. In a similar way, the covariance between $y(s)$ and $y(s + h)$ can be computed as

$$C(y(s), y(s+h)) = \frac{1}{n(h)} \sum(y(s) - \bar{y}(s))(y(s+h) - \bar{y}(s+h)), \tag{2.48}$$

where $n(h)$ is the number of paired comparisons at lag h. The covariance can be computed for different lags of h and the plot of covariance as a function of lags(h) is called an autocovariance diagram.

Semivariogram

The degree of spread about a line is known as the moment of inertia and is calculated as γ=$\frac{1}{2n}\sum(x_i - y_i)^2$. Similarly, the moment of inertia about a support is called semivariogram and it is expressed as

$$\gamma(h) = \frac{1}{2n(h)}\sum(y(s) - y(s+h))^2 \qquad (2.49)$$

This method of computing semivariances is often called the Matheron's method of moments (MoM) estimator. More details on this method and other methods of estimating semivariograms can be found in Fischer & Getis (2010).

2.3.6.3 Estimation of spatial dependence

There exists a clear distinction between spatial variation of the risk and spatial clustering (Bivand et al., 2008). Spatial variation is a phenomenon that occurs when the risk is not homogeneous in the domain of study. This means that the likelihood of contracting a certain disease is not the same for all members of the population at risk even though cases are assumed to be independent of each other vis a vis the to underlying risk. On the other hand, the spatial clustering phenomenon assumes that risk occurrence is homogeneous through the entire study region and the presence of a case affects the odds of risk of other individuals in its neighbourhood.

Method of assessing spatial autocorrelation: Semivariogram

Spatial dependence is one of the vital characteristics of spatial data. It shows how observations that are close together tend to be more similar than those farther apart spatially. Like in typical statistical methods where the correlation may be computed by means of a scatterplot for a number of data points (x, y), the spatial association between values of a variable $y(s)$ can be estimated using a semivariogram. One way to explore the spatial correlation is to use the variogram cloud, which is obtained by plotting all possible squared differences of observation pairs $(y(s_i) - y(s_j))^2$ against their separation distance h_{ij}. However, to estimate the spatial correlation from observational data, two important assumptions are usually made: intrinsic stationarity and isotropy of the spatial process $y(s)$.

Under these assumptions, the variogram of a spatial stochastic process $y(s)$ is the function

$$Var\{y(s_i) - y(s_j)\} = Var(y(s_i)) + Var(y(s_j)) - 2Cov(y(s_i) - y(s_j)) = \sigma^2(1 - \rho) \tag{2.50}$$

We note that the half of the variogram is a semivariaogram. The following are some important properties of semivariogram to qualify as a valid one.

- $\gamma(y(s), y(s+h))$=$\gamma(y(s+h), y(s)$, i.e. the spatial correlation between y(s)and y(s+h) is the same as the spatial correlation between y(s+h) and y(s).

- $\gamma(h = 0)$=0 as $Var(y(s) - y(s)) = 0$

- $\frac{\gamma(h)}{\|h\|^2} \to 0$ as $h \to \infty$

- $\gamma(\cdot)$ must be conditionally negative definite. That is, for any finite number of locations $s_i : i = 1,\ldots,m$ and real numbers $a_1,\ldots,$ a_m, then $\sum_{i=1}^{m} \sum_{j=1}^{m} a_i\ a_j\ \gamma(s_i, s_j) \leq 0$.

A graph of a semivariogram plotted against separation distance ($\| h \|$) conveys information about the continuity and spatial variability of the stochastic process. If near observations are more similar than those located farther apart, then the graph may start at zero and gradually increases until it reaches a constant value, referred to as a sill, as the separation distance increases. This separation distance at which the graph reaches its apex is called the range. This indicates that the spatial autocorrelation increases with decreasing separation distance within the range. In case the graph does not start from zero, it implies that there is a discontinuity at the origin and the spatial process has a nugget effect. Although the nugget effect is quite often signaling some measurement error variance (τ^2) in the spatial process, it may also be an indication of a natural discontinuity in the spatially process. If the process has a large nugget effect, it is possible for two locations fairly close together to have very different values. A possible example is found in gold mining sector where ore may not be found at one location, but then at a nearby location a mass of gold (i.e. gold nugget) is found. In the field of geostatistics, a number of semivariogram models have been used, of which the most popular are given below.

Nevertheless, a comprehensive review of other semivariogram models can be found in Waller & Gotway (2004).

Exponential semivariogram model

An exponential semivariogram is defined as follows

$$\gamma(h,\theta) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_e\{1 - \exp(-\frac{\|h\|}{a_e})\} & \text{if } h \neq 0 \end{cases} \tag{2.51}$$

where $\theta = (c_0, c_e, \|h\|, a_e)^t$, $c_0 \geq 0$ is the nugget, $c_e \geq 0$ is the partial sill, and $c_0 + c_e$ is the sill, $\|h\|$ is the separation distance, and the effective range, which is conventionally defined as the distance at which the autocorrelation equals 0.05, is $3a_e$.

Gaussian semivariogram model

The function of a Gaussian semivariogram model is expressed as

$$\gamma(h,\theta) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_g\{1 - \exp(-(\frac{\|h\|}{a_g})^2)\} & \text{if } h \neq 0 \end{cases} \tag{2.52}$$

where $\theta = (c_0, c_g, \|h\|, a_g)^t$, $c_0 \geq 0$ is the nugget, $c_g \geq 0$ is the partial sill, and $c_0 + c_e$ is the sill, $\|h\|$ is the separation distance, $a_g \geq 0$, and the effective range $\sqrt{3}a_e$.

K-Bessel (Mat$\acute{e}$rn) semivariogram model

A K-Bessel semivariogram model is mathematically defined as follows

$$\gamma(h,\theta) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_k\{1 - \frac{1}{2^{\alpha-1)}\Gamma(\alpha)}(\frac{\|h\|}{a_k})^\alpha K_\alpha \frac{\|h\|}{a_k}\} & \text{if } h > 0 \end{cases} \tag{2.53}$$

where $\theta = (c_0, c_k, \|h\|, a_k, \alpha)^t$, $c_0 \geq 0$ is the nugget, $c_k \geq 0$ is the partial sill, and $c_0 + c_k$ is the sill, $\|h\|$ is the separation distance, $a_k \geq 0$, $K_\alpha(\cdot)$ is the modified Bessel function of the second kind of order α, and $\Gamma(\cdot)$ is the gamma function. The sill is approached as the separation distance tends to infinity.

Power semivariogram model

Mathematically, a power semivariogram model is described by

$$\gamma(h,\theta) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + b\|h\|^p & \text{if } h \neq 0 \end{cases} \tag{2.54}$$

where $\theta = (c_0, b, \|h\|, p)^t$, $c_0 \geq 0$ is the nugget, $b \geq 0$, and $0 \leq p \leq 2$, and $\|h\|$ is the separation distance. This family of models does not have a sill nor a range. That is the spatial correlation does not decrease as lag distance increases. If $p = 1$, the model becomes linear.

Spherical semivariogram model

The function of a spherical semivariogram model is defined as

$$\gamma(h,\theta) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_s\{\frac{3}{2}(\frac{\|h\|}{a_s}) - \frac{2}{2}(\frac{\|h\|}{a_s})^3\} & \text{if } 0 \leq h \leq a_s \\ c_0 + c_s & \text{if } h > a_s \end{cases} \tag{2.55}$$

where $\theta = (c_0, c_s, \|h\|, a_s)^t$, $c_0 \geq 0$ is the nugget, $c_s \geq 0$ is the partial sill, $c_0 + c_k$ is the sill, $\|h\|$ is the separation distance, and $a_s \geq 0$ is the range. Unlike for the other models discussed earlier, which are valid in one or more dimension space, the spherical model is valid only in $\mathbb{R}^d$, $d = 1, 2, 3$.

2.3.6.4 Spatial interpolation and prediction methods

Waller & Gotway (2004) define the interpolation as the process of obtaining a value for a Gaussian stochastic process of interest at an unsampled location (usually denoted as $Z(s_0)$) based on measurements taken from surrounding locations. Interpolation techniques are classified into two broad categories, namely deterministic and stochastic (probabilistic) interpolators. The main difference between the two classes is that the latter has a probability model for data and thus provides statistics such as standard errors while the former does not assume any probability model. When the latter method is used for interpolation, it is known as a method for spatial prediction.

Deterministic interpolation techniques

There exists quite a number of deterministic interpolation methods that include inverse distance weighted, global polynomial, local polynomial, and radial basis functions. Due to their weakness of not providing standard errors of interpolated values, these models are rarely used. Here we only review the inverse distance weighted (IDW) method. This interpolation method has been commonly used among other deterministic models and it obeys Tobler's law of geography as it assumes that each input point has a local influence that decreases with distance (Naish, 2012). This technique is based on a simple weighted average of neighbouring values where the resulting interpolating surface should be hugely depending on nearby values but not on values farther apart. The mathematical expression of a general inverse-distance interpolator is given by

$$\hat{Z}_0 = \frac{\sum_{i=1}^{n} \parallel s_i - s_0 \parallel^{-p} Z(s_i)}{\sum_{i=1}^{n} \parallel s_i - s_0 \parallel^{-p}} \tag{2.56}$$

where $\hat{Z}_0$ is the interpolated value, $\parallel s_i - s_0 \parallel$ is the euclidean distance between i^{th} location (s_i) and the unsample location s_0, and $1 \leq p \leq 3$ is the power. When $p = 2$, the technique is known as inverse-distance-squared interpolator, which is the most popular interpolation method (Waller & Gotway, 2004). The popularity of the inverse distance interpolation is due to its mathematical simplicity and flexibility with respect to computation. We note that the weight assigned to each observation is an inverse function of the distance between that observation's location and the unsample location s_0 at which interpolation is needed. If care is taken to ensure that the interpolated values are based on enough data and a right power, the IDW method is known to yield a fairly accurate value.

Spatial prediction methods

Originally, geostatistcs modelling was developed to predict the probability distribution of ore grades in mining industry. In 1950, D.G. Krige developed an interpolation technique, named after his name as the kriging method, for use in the South African mining industry (Gelfand et al., 2010). In collaboration with G Marheron, a French mathematician at the Ecole of Mines, this method was further improved

(Van Beers & Kleijnen, 2003). Due to this background, quite often, the geostatistics modelling was regarded as a statistical modelling technique applied only to geological data. However, nowdays, kriging and its derivatives are enjoying a wide range of applications in diverse disciplines.

Kriging methods are preferred over other interpolation methods as they provide the best linear unbiased estimators (BLUEs) (Isaaks & Srivastava, 1989; Van Beers & Kleijnen, 2003). Kriging techniques include among others simple kriging, ordinary kriging, universal kriging, factorial kriging, block kriging, indicator kriging, stratified kriging, Poisson kriging and cokriging. Despite the existence of many types of kriging methods, the present review explores only simple kriging. The interested reader may consult other sources (e.g. Isaaks & Srivastava (1989); Waller & Gotway (2004)) for a detailed review of other types.

Let $Z(s)$ be the regionalised random variable, where $Z(s_i)$ refers to the measurement of Z obtained at point location s_i, and $Z(s_0)$ is assigned to the location where the regionalised variable is to be estimated. Then, the spatial variation of $Z(s)$ consists of a trend component (mean) $m(s)$, a spatial stochastic component $R(s)$, and a random Gaussian noise (ε) with mean zero and constant variance σ^2. Mathematically, it is expressed as

$$Z(s) = m(s) + R(s) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \tag{2.57}$$

where $m(s) \in \mathbb{R}$ and $R(s)$ is a zero-mean intrinsically stationary random process with variogram $2\gamma(\cdot)$. Assuming the stationarity, the interpolated value at unsampled location s_0 is generally expressed as

$$\hat{Z}(s_0) = m(s_0) + \sum_{i=1}^{N(h)} \alpha_i (Z(s_i) - m(s_i)) \tag{2.58}$$

The primary objective of kriging methods is to provide best linear unbiased estimators at unsampled locations s_0 based on weighted average of adjacent locations within a given search area and the sum of weights must be equal to one in order to ensure that estimates are unbiased. The difference between simple and ordinary Kriging methods is fundamentally based on the assumption made about the functional form of $m(s)$. For simple kriging, $m(s)$ is assumed to be a known constant whereas for ordinary kriging, $m(s)$ is assumed to be unknown but constant.

Simple kriging

Since the component trend $m(s)$ is known and constant, say $m(s)$=$m(s)$=μ, then Eq. (2.58) becomes

$$\hat{Z}(s_0) = \mu + \sum_{i=1}^{N(h)} \alpha_i (Z(s_i) - \mu) \tag{2.59}$$

With little algebraic manipulation and taking into that consideration the estiamators are unbiased (i.e. $\sum_{i=1}^{N(h)} \alpha_i = 1$), the Eq. (2.59) simplifies to

$$\hat{Z}(s_0) = \sum_{i=1}^{N(h)} \alpha_i Z(s_i) \tag{2.60}$$

The simple kriging weights are obtained by minimizing the estimate of the error variance $\sigma_E(s_0)$. Through some algebraic manipulation, the weights α_i are obtained using α_i=$K^{-1}k$, where

$$\mathbf{K} = \begin{pmatrix} C(s_1 - s_1) & C(s_1 - s_2) & \dots & C(s_1 - s_{N(h)}) \\ C(s_2 - s_1) & C(s_2 - s_2) & \dots & C(s_2 - s_{N(h)}) \\ \vdots & \vdots & \dots \quad \vdots & \\ C(s_{N(h)} - s_1) & C(s_{N(h)} - s_2) & \dots & C(s_{N(h)} - s_{N(h)}) \end{pmatrix}$$

is a matrix of data covariances and

$$\mathbf{k} = \begin{pmatrix} C(s_1 - s_0) \\ C(s_2 - s_0) \\ \vdots \\ C(s_{N(h)} - s_0)) \end{pmatrix}$$

is a matrix of covariances between the observed data and the unsampled location, and

$$\alpha_{\mathbf{i}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{N(h)} \end{pmatrix}$$

In a practical situation, it is not always possible to know the trend component $m(s_0)$ at an unsampled location, hence, simple kriging is less commonly used.

Because ordinary kriging assumes a constant unknown trend component, it is by far the most widely used type of kriging relative to others.

2.3.6.5 Spatial modelling of geostatics data using the SPDE approach

In classical geostatistics as well as in modern hierarchical spatial modelling, Gaussian fields (GFs) are considered as a corner stone of the modelling aspect. GFs are both analytically and practically convenient because they possess explicit and computable normalising constants and good analytic properties. Quite often, the specification of a Gaussian field is achieved through a mean function ($\mu(\cdot)$) and a covariance matrix (Σ) whose elements are function of a covariance function and usually the Mat$\acute{e}$rn covariance function is assumed. This covariance matrix is in most cases dense, which results in computational issues known as the big n problem.

To overcome the computation issues caused by the big n problem associated with dense matrix, Lindgren & Rue (2011) suggested a method that makes use of the fact that a Gaussian field with a Mat$\acute{e}$rn covariance function is a solution to a certain linear fractional partial differential equation (SPDE). This method consists of the following main steps. First, find a GMRF that best represents the GF. That is, the GMRF should have a local neighbourhood and a precision matrix whose inverse Q^{-1} is close to the covariance matrix of the GF. Second, the computations are done using the GRMF representations through the use of a set of spatial random functions with weighted sums of simple basis functions in order to conserve the continuous interpretation of a GF. Explicitly, a GRMF representation is constructed by using a stochastic partial differential equation (shown in Eq. (2.61)) which has GFs with the Mat$\acute{e}$rn covariance function as a solution.

$$(\kappa^2 - \triangle)^{\frac{\alpha}{2}} x(u) = w(u), \quad u \in \Re^d, \quad \alpha = \nu + \frac{d}{2}, \quad \kappa > 0, \quad \nu > 0, \tag{2.61}$$

where $(\kappa^2 - \triangle)^{\frac{\alpha}{2}}$ is a pseudo differential operator, $\triangle {=} \sum_i^d \frac{\partial^2}{\partial x_i^2}$ is the Laplacian, κ is the spatial scale parameter, α is the smoothness parameter, $x(u)$ is a GF, and $w(u)$ is the spatial white noise. The solution of the SPDE in Eq. (2.61) is called a Mat$\acute{e}$rn field. Integer values for α give continuous markov fields from which discrete basis representations can be obtained (Lindgren, 2012).

Two main results of Eq. (2.61) are presented in Lindgren & Rue (2011).
With the first result, it has been shown that an approximately weak solution to stochastic partial differential equations can be used to provide an explicit link, which is expressed as a basis function representation, between some GFs in the Mat$\acute{e}$rn family and GRMFs for any triangulation on a regular grid of $\mathbb{R}^d$.

The second result, which is an extension on irregular grids, is reviewed below. For $d = 2$, the domain $\mathbb{R}$ is subdivided into a collection of non-intersecting triangles with the condition that any two triangles meet in at most a common edge or corner. Generally, initial vertices are placed at the locations for observations and then additional vertices are added in a way that minimises the number of triangles needed to fill up the size and shape of the study domain of interest. This results to a constrained refined Delaunay triangulation, also known as a mesh. Once a stochastic weak solution formulation to the SPDE is found, a construction of a finite element representation of this solution is obtained using

$$x(u) = \sum_{i=1}^{n} \psi_i(u) w_i, \tag{2.62}$$

where n is the number of vertices in the triangulation, $\{w_i\}$ are Gaussian distributed weights, and $\{\psi_i\}$ are basis functions, usually a piecewise linear function in each triangle. At the vertex i, ψ_i takes the value 1 and zero at all other vertices. The joint distribution of w=$\{w_i, \ldots, w_n\}$ is selected in a way that the distribution of functions $x(u)$ approximates the distribution of solutions to Eq. (2.61). This triangulation process results in a diagonal matrix C and a sparse matrix G, such that the precision matrix for the weights is expressed as (Lindgren, 2012)
$Q \approx \kappa^4 C + 2\kappa^2 G + GC^{-1}G, \quad C_{ij}= \langle \psi_i, \psi_j \rangle$ and $G_{ij}= \langle \bigtriangledown\psi_i, \bigtriangledown\psi_i \rangle$

While constructing a triangulated mesh on top of which the SPDE/GMRF representation is to be built, boundary effects from SPDE need to be taken care of. In case of a stationary field across the entire domain of observations, it is suggested that the model domain should be extended far enough in order to avoid the influence of boundary effects on observations (Lindgren, 2012). This can be achieved by creating a polygon of triangles out of the domain area known as a convex hull (Krainski & Lindgren, 2013).

Recall that the SPDE model is defined at the mesh vertices (m dimension) and the vertices are not necessarily defined at n location points where the response variable is observed. Thus, there is need to specify how the Gaussian Markov random field and other model components are linked to the response. The specification is achieved through a projector matrix that projects the process at the mesh vertices to the locations response.

A simple SPDE model is then defined using the *inla.spde2.matern(·)* object where the prior distributions of parameters are defined according to a specific situation. The Bayesian inference is based on the integrated nested Laplace approximation. More details on fitting a SPDE model can be obtained from (Krainski & Lindgren, 2013).

2.3.7 Spatio-temporal modelling

2.3.7.1 Introduction

In previous sections, the reviewed analysis methods focused on the spatial aspect of data, but they did not account for the temporal aspect that might be present in the actual data. This means that the stochastic process $\{Y(s) : s \in D \subseteq \mathbb{R}^d\}$ was assumed to vary only as a function of the spatial location s. To include the temporal aspect, we now consider processes, $\{Y(s,t) : (s,t) \in D \subseteq \mathbb{R}^d \times \mathbb{R}\}$, which are a function of both the spatial location, $s \in \mathbb{R}^d$, and time, $t \in \mathbb{R}$. This yields a spatio-temporal process. Although some of the concepts in analysis of spatio-temporal observations are accepted as generalisations of those developed for spatial data, it is worthy to note that time differs intrinsically from space. For instance, time moves only forward, while there might be many directions in space. Thus, it is easy to define temporal lags, whereas it is difficult to define spatial lags that are comparable with temporal lags.

Let $y(s,t)$ =$m(s,t) + R(s,t) + \varepsilon(s,t)$ be the Gaussian process, which is an extension of the spatial process presented in Eq. (2.57). Here $m(s,t)$ represents a deterministic space-time trend function, $R(s,t)$ is a stationary process with mean zero and continuous sample paths, and $\varepsilon(s,t)$ is an error field, also called the nugget effect, with mean zero and discontinuous realizations, which is independent of R. The main objective of spatio-temporal modelling is to specify the mean structure and covariance structure, and interpolate the value of $y(s_o, t_0)$ at an unsampled location (s_0) at a specific point in time (t_0). In what follows, we first review the mean structure modelling and thereafter we move on with the review of covariance structure modelling.

2.3.7.2 Mean structure models

The space-time trend function, $m(s,t)$, consists of a purely spatial component, a purely temporal trend component, and a space-time interaction component. For count data, the space-time trend component may generally be expressed as follows

$$log(m(s,t)) = m_0 + A_s + B_t + C_{st}, \tag{2.63}$$

where $s = 1, \dots, n$, $t = 1, \dots, T$, A_s is a spatial component, B_t is a temporal component, and C_{st} is a space-temporal component. In the literature, there exist many parametrizations of Eq.(2.63). For example, Bernardinelli et al. (1995) formulated the linear predictor as

$$log(m(s,t)) = m_0 + u_s + \nu_s + (\beta + \delta_s) \times t \tag{2.64}$$

The parametric trend for the temporal component consists of the main linear trend β which is the representation of the global time effect, and δ_s is the interaction between time and space. u_s assumes a CAR prior distribution, that is $u_s \mid u_{-s} \sim N(\frac{1}{n_s} \sum u_s, \frac{1}{\tau_{n_s}})$; ν assumes exchangeable prior (i.e. $\nu_s \sim N(0, \sigma^2_{\nu_s} \ )$; β assumes a weak informative Gaussian prior $\beta \sim N(0, \sigma^2_\beta)$; and δ_s may assume a Gaussian prior or CAR structure. Another formulation of the mean structure, which relaxes the restrictive linearity assumption imposed on the differential trend δ_s, which is expressed as follows (Knorr-Held, 1999)

$$log(m(s,t)) = m_0 + u_s + \nu_s + \gamma_t + \phi_{st} \tag{2.65}$$

Conditional autoregressive prior distributions are quite often assumed for γ_t. That is $\gamma_t \sim N(\xi \gamma_{t-1}, \frac{1}{\tau_{\gamma_t}})$. If $\xi = 1$, then γ_t becomes a non-parametric random effect with a random walk prior. ϕ_{st} is commonly assumed to follow a Gaussian prior with mean zero (i.e. $\phi_{st} \sim N(0, \sigma^2_\phi)$). ϕ_{st} may assume other types of prior distributions such as type I interaction, type II random walk interaction, type III interaction consisting of time-averaged spatial correlation, and type IV interaction, which is fully space-time dependent (see Lawson (2013); Blangiardo & Cameletti (2015)). Often, temporal trends are periodic and exhibit some seasonal effects and hence can be modelled with trigonometric functions or seasonal models. Additional to the spatial and temporal coordinates, the trend component might be a function of environmental temporal and/or spatial covariates such as temperature or population density.

2.3.7.3 Covariance structure models

Unlike for spatial domain, the estimation of space-time covariance structures can be problematic if no bridging assumptions are made. For instance, under the second order stationary assumption of a stochastic process, by definition, in spatial setting the covariance is given by $C(h) = C(-h)$. But, in the spatio-temporal context, $C(h, u) = C(-h, u) = C(h, -u)$ does not hold each time (h and u are spatial and temporal shifts, respectively). The following are some important properties of a space-time covariance functions:

- *Separable* space-time covariance function:

A space-time covariance function is defined to be separable if it decomposes into two components, namely the purely spatial component and the purely temporal component. That is $C(h, u) = C(h)C(u)$. While this assumption has numerous advantages (e.g. parsimony of the model and enhances fast computation for large space-time data), this assumption does not cater for space-time interactions. In addition, separable covariance models fail to fit most physical situations as they are too simplistic (Gelfand et al., 2010). A separable space-time covariance function can be constructed by multiplying together any valid spatial covariance function and valid temporal covariance function. A simple example is using a spatial correlation from a member of the exponential family together with a temporal correlation from an autoregressive process of order 1 (AR(1)) with parameter $\rho = \exp(-\nu_t)$ to produce a separable spacetime covariance (Sherman, 2011) expressed as
$C(h, u, \theta)= \sigma^2 \exp(-\nu_s \parallel h \parallel) \exp(-\nu_t \mid u \mid)$, where $\theta = (\sigma^2, \nu_s, \nu_t)$, σ^2 is a scale parameter, and ν_s and ν_t are decay parameters.

- *Fully symmetric* space-time covariance function:

A space-time covariance function is fully symmetric if for any two locations, the model is unable to distinguish possible differing effects as time moves forward or backward. That is expressed mathematically as
$C(h, u) = C(-h, u) = C(h, -u)$.

- *Nonseparable* space-time covariance function:

As separable covariance models have frequently been unable to fit physical phenomena and observational data, a class of nonseparable functions has been sought. Various scholars have played important role in developing nonseparable space-time covariance functions. These functions have been constructed through partial differential equations and through spectral densities. In addition, general classes of nonseprable covariance functions were constructed using closed Fourier inversion form in $\mathbb{R}^d$ (Sherman, 2011). Gneiting (2002) introduced a Fourier-free implementation of nonseparable and stationary covariance functions that allow for space-time interactions and expanded the class of valid space-time covariance functions. An easily interpretable nonseparable space-time covariance function with interaction parameter (Gneiting, 2002) is expressed as

$$C(h, u, \theta) = \frac{\sigma^2}{(\mid u \mid^{2\gamma} + 1)^{\tau}} \exp\{\frac{-c \parallel h \parallel^{2\gamma}}{(\mid u \mid^{2\gamma} + 1)^{\beta\gamma}}\}, \tag{2.66}$$

where $\theta = (\beta, \gamma, \sigma^2, \tau)$, $\beta \in (0, 1]$ quantifies the strength of the space-time interaction, $\gamma \in (0, 1]$ is a smoothness parameter of the spatial correlation, τ is a smoothness parameter of the temporal correlation, and c indicates the strength of the spatial correlation.

2.4 Some current issues in spatial and spatio-temporal modelling

2.4.1 Misalignment

In epidemiology, ecology, agriculture and geology, and other many fields, relating data collected at different scales, locations and dimensions poses challenges in spatial analysis. With the increasing availability of geographically referenced data, linking of collected data is indeed unavoidable as the exploitation of this readily available information helps avoiding the implementation of new and expensive data collection. In statistical literature, the analysis of originally collected data on one resolution with the purpose to make inferences on a different level of spatial resolution is referred to as the misalignment problem (Gotway & Young, 2002). This problem may present itself in many different facets with varying characteristics.

Recent advances in geographic information systems (GIS) and internet make it possible to access spatial data in various forms that include point, line, area, surface, etc. But one major concern is how best these data can be integrated to answer real life problems. The integration of such information may require the data transformation as the spatial process of interest intrinsically present in one form of data may completely be different from the one observed in another form of data. In spatial statistics, one commonly used transformation is the change of support. In statistical terms, a support means the shape, size, and volume associated with each data value, and which also extends to the spatial orientation of the domains of study associated with each spatial measurement. Thus, changing the support of the spatial random process generates a new variable related to the old one but with new statistical properties. Consequently, every time a support is changed, it implies that new statistical properties of a spatial process should be studied. In statistical literature, this problem is referred to as a change of support problem (COSP)(Waller & Gotway, 2004). Table 2.3, adapted from Gotway & Young (2002), shows some of the commonly encountered change of support problems in spatial modelling. A discussion of some of the change of support problems is given in the subsequent paragraphs.

Modifiable Areal Unit Problem (MAUP) and Ecological fallacy (EF) are examples

Table 2.3: Examples of change of support problems (Gotway & Young (2002))

Process observed at spatial level	Inference at spatial level
Point	Point
Area	Point
Point	Line
Point	Area
Point	Surface
Area	Area

of COSPs. These problems are caused by aggregation. In general, aggregation reduces heterogeneity among units or individuals as the uniqueness of each unit or individual is lost.

A Modifiable Areal Unit Problem arises when one wishes to use a variable observed at the areal-level of spatial aggregation in order to make inference at another areal-level of aggregation. Stated differently, this problem refers to inference made using spatial data at a different level of spatial resolution than it was originally collected. For example, a spatial random process Y is observed at k blocks (i.e. $Y(B_1), \ldots, Y(B_k)$) but predictions $Y(\acute{B}_1), \ldots, Y(\acute{B}_k)$ are to be made from observed blocks data. Specifically, one may wish to estimate the relative risk of a given disease at enumeration area level (lower level) using counts observed at constituency level (higher level). MAUP leads to two side effects namely scale and group effects. The scale effect, also known as the aggregation effect, refers to obtaining different inferences as aggregation into increasingly larger areal units is made. The zoning effect also referred to as grouping effect concerns the variability in results due to differences in shape of areal units even if they are at the same scale of aggregation.

Ecological fallacy occurs when inferences made using aggregated data may not accurately reveal the same inferences as those ones would be obtained if individual level data were to be used. It generally refers to as the inference about the point level made from the aggregate level. For example, a spatial random process Y is observed at a finite blocks (Bs) but the inference is about $Y(\acute{s}_1), \ldots, Y(\acute{s}_k)$, starting from $Y(B_1), \ldots, Y(B_k)$. Quite often, associations observed between variables measured at aggregate level overstate the relationships in the same variables when measured at the individual level. In other words, using aggregated data to infer at individual level results in biased conclusions. In the literature, the resulting bias is referred to as the ecological bias, which comprises of aggregation and specification biases (Gotway & Young, 2002). These effects are similar to the aggregation effect and the zoning effect discussed under MAUP.

In spatial regression modelling, it is quite common to encounter situations whereby the dependent (Y) variable and its explanatory variable (X) are spatially misaligned. This type of misalignment may be manifested in one of the following facets.

- Point to point misalignment:

X is at one point level and Y is observed at a different point level. For example, the purpose is to relate an exposure, say pollution observed at specific sites, to a chronic respiratory disease observed at an individual level.

- Point to area:

An explanatory random variable X , available at point level , is to be related to Y, which is observed at areal level. Also, the points to points misalignment problem may occur if for instance a spatial random process $Y(S)$ is observed at a finite set of sites, s_i, $i = 1, \ldots, k$, (i.e. $Y(s_1), \ldots, Y(s_k)$) and the interest is to predict $Y(\acute{B}_1), \ldots, Y(\acute{B}_k)$, where $\acute{B}_i$s are blocks (or areal units) in the study domain.

- Block to bloc misalignment:

Both X and Y are available at different levels of aggregation.

Another form of misalignment may occur if one wishes to integrate two data sets measured at different levels of aggregation. This may be the case of jointly modelling two more diseases.

Various methods of resolving misalignment, which include kriging method, Monte carlo integration, downscaling methods and Bayesian models, have been proposed (Goovaerts, 2008; Keil, Belmaker, Wilson, Unitt, & Jetz, 2013; Sturrock et al., 2014; Araújo, Thuiller, Williams, & Reginster, 2005; Lee & Sarran, 2015; Finley, Banerjee, & Cook, 2014; Illian, Møller, & Waagepetersen, 2009). For instance, methods have been applied to downscale the distributions of data from coarse to fine grain that include direct method, point sampling method, and hierarchical Bayesian method, which have all been adopted to deal with this scenario of misalignment (Keil et al., 2013; Sturrock et al., 2014; Araújo et al., 2005). Other techniques have been developed that deal with the spatial misalignment that arises when the response variable is available at bigger and irregular shaped area units, and where covariates are available at smaller fine grids (Lee & Sarran, 2015). In the case where misalignment occurs with non-nested overlapping grids, hierarchical Bayesian approaches have been employed (Banerjee et al., 2004; Finley et al., 2014). Of recent, the latter has been extensively applied as it permits to derive posterior predictive distributions for both parameters and it enables to incorporate additional sources of information in a form of prior knowledge in order to deal with multiple sources of uncertainty (Illian et al., 2009).

For jointly modelling different data sets, different approaches of multivariate techniques that include the multivariate normal distribution, iterative generalised least squares (IGLS) method, multivariate conditional autoregressive (MCAR) modelling, and the shared-component modelling are commonly used in the spatial analysis of multiple diseases (Manda, Feltbower, & Gilthorpe, 2012). A reparametrised and marginalised posterior sampling (RAMPS) algorithm was introduced by (Yan, Cowles, Wang, & Armonstrong, 2007) with the purpose of lowering autocorrelation in MCMC samples, which is known to lead to computational convergence problems when anlaysing large spatiotemporal data sets. Cowles, Yan, & Smith (2009) further extended RAMPS to allow jointly modelling of areal and point-referenced data. An illustration of the implementation of RAMPS algorithm in the R package ramps can be found in Smith, Yan, Cowles, et al. (2008).

For point referenced data, an adaptive geostatistical sampling technique has been perceived as a tool that helps avoiding to report results derived from multiple data sources which are quite often known to have different accuracies and to be spatially and temporally misaligned (Kabaghe et al., 2017). This type of sampling technique replaces the process of sampling in a single phase by splitting it up into several successive phases. On any sampling phase, the choice of sampling units is informed by the results computed from information obtained from the previous sampling phase (Kabaghe et al., 2017). In other words, it enables the collection of both response variables and its covariates to depend on the information previously gathered (Chipeta, Terlouw, Phiri, & Diggle, 2015). Although this sampling method yields better representative surveys relative to surveys resulting from the traditional ways of sampling, it avoids integrating information readily available from the multiple sources.

2.4.2 Edge effects

Borders of a domain of study or physical barriers such as rivers or a forest may define the boundaries or edges of a study area. In most of cases, the area beyond the edges may have incomplete data or no data at all. Also, areas at boundaries have very few neighbours relative to those in the center of the study domain as the areas beyond the edges are not part of the study region. Thus, any spatial analysis based on borrowing strength from neighbouring areas may produce distorted results at points or areas at edges since very few neighbours are available. In the literature, these distortions are generally referred to as boundary or edge effects. The boundary effects constitute a major problem in smoothing because of the nonavailability of data or fewer data that are available near the boundaries, and as well due to the properties of the particular smoothing technique being employed. For example, when using Ripley K-function to analyse a point pattern at a range of scales and to determine at which scales these points tend to be regular or clumped, the edge effect occurs within the search circle.

Many methods of dealing with edge effects have been proposed with weighting systems and guard areas being commonly used. Weighting systems are based on setting up weights that relate the position of the point or area to the external edge. This approach gives less weights to observations near the edges. The weights assigned to observations act as proxies of the degree of missing information at the specific locations (Lawson et al., 1999). For situations whereby a small portion of study area is near the boundaries and the purpose is to estimate the overall parameter, then the edge correction method based on weighting may be used to attenuate the boundary effects. In hierarchical Bayesian modelling, it is proposed to accommodate edge-weighted data by weighing each area through the addition of an offset term in the linear predictor of the regression model (Lawson et al., 1999).

Another way of dealing with edge effects is to employ external areas to the main domain of study. These areas are known as guard areas and they can be constructed by adding an area to the study window or by considering some fixed distances from the external boundaries. For both point and areal spatial processes, it is recommended to use internal or external guard areas through augmentation achieved by Monte Carlo Markov Chain (MCMC) simulations. Although the guard areas are used in the estimation process, the results at these areas are not reported because they are subjected to boundary effects themselves (Lawson et al., 1999).

Another alternative way of handling edge effects is to consider that data are missing along the boundaries of the study area and estimate the missing data (Griffith, 1985). Other edge correction methods proposed by Griffith (1985) are based upon a dummy variable discrimination between border areal units and non-edge areas and on a generalised least squares (GLS) type of adjustment matrix. For the point pattern spatial process, boundary correction methods for K-function have been proposed: Reply's circumference correction, toroidal correction and guard area correction (Yamada & Rogerson, 2003).

2.4.3 Measurement error models

In almost every statistical analysis, it is assumed that all observations obtained from variables involved in modelling are error free. But, in many situations, this assumption is not met. For instance, the variable of interest cannot be measured correctly. This may occur when a researcher is interested in measuring the average sugar intake, intelligence, or age. Because the true value is not measurable, a surrogate measurement is taken instead. The measurement error is defined as the discrepancy between the true value and the observed value of random variable.

Quite often, one hopes to have accounted for measurement errors by including a disturbance term, sometimes called an error term (ε), in the model. This term is generally intended to represent unexplained variability due to explanatory variables that might have been actually excluded in the model. It has been argued that as long as measurement errors are negligible in magnitude, then they can be assumed to be merged in the disturbance term and they will have minimal effects on the statistical inferences. Otherwise, they will affect statistical inferences (Chen, Hong, & Nekipelov, 2007).

2.4.3.1 Types of measurement errors

- **Classical measurement errors**

Assume a variable y is linearly dependent on x. Further, assume that y is error free whereas x is observed with errors in additive form. That is $w = x + \nu_x$, where ν_x is error associated with x and it has a zero mean and variance σ_x^2. Thus, a simple linear model is given by

$$y = \beta x + \varepsilon \tag{2.67}$$

The measurement error in the explanatory variable (ν_x) is said to be a classical measurement error if it has a zero mean, is uncorrelated with the true dependent (y) and (x), and is uncorrelated with the disturbance term (ε) (Pischke, 2007). That is, $E(\nu_x) = 0$, $plim\frac{1}{n}(\acute{y}\nu_x) = 0$, $plim\frac{1}{n}(\acute{x}\nu_x) = 0$, and $plim\frac{1}{n}(\acute{\varepsilon}\nu_x) = 0$. With a little algebraic manipulation of Eq. (2.67), it can be shown that the error in x becomes part of the disturbance term and thus creating a bias (i.e. $y = \beta w + (\varepsilon - \beta\nu_x)$.

- **Non-classical measurement errors**

If it is assumed that error term in ν_x is correlated with the explanatory variable x, then it is said to be a non-classical error. One example of this type of error is the one generated by a misclassification of the binary regressor. In this case, the measurement error is negatively correlated to the true dichotomous variable.

2.4.3.2 Measurement error models

The main aim of measurement error modelling is to obtain nearly unbiased estimates of the parameters by indirectly fitting a model for Y in terms of observed explanatory covariate (W) prone to errors. However, Carroll, Ruppert, Stefanski, & Crainiceanu (2006) warned that a direct substitution of W for X with no adjustment in the normal routine of model fitting may lead to biased estimators. In addition, the mis-specification of a measurement error model leads to erroneous inferences. The specification of a measurement error model is based on an assumption about the distribution of the observed values given the true values or vice versa (Buonaccorsi, 2010). For the classical measurement error model, the distribution of the observed values given the true values is specified, while the latter specification is referred to as the Berkson error model. That is, the classical measurement error model is expressed as $P(W = w \mid x)$, while the Berkson error model is given by $P(X = x \mid w)$. X and W are the true and observed covariates, respectively.

- **Additive non-differential measurement error model**

An additive non-differential classical measurement error model assumes that the measurement error does not depend on the value of the response and $w \mid x = x + \nu$. In this case, w are observed values of the true but unobserved covariates X (i.e. Ws are surrogate of Xs). The error term ν can assume a Gaussian prior with a zero mean and a covariance matrix $C = \tau_\nu D$ (i.e. $\nu \sim N(0, C)$), where τ_ν is the precision of the error term and D is the diagonal matrix of fixed scaling values (d_i) of the observational precision.

The Berkson error model specifies the distribution of $X \mid W$. Let us assume that the response variable is measured without error and only X is measured with error.

The additive Berkson model with a constant variance assumes that
$X = W + e_i$ with $E(e_i) = 0$ and $Var(e_i) = \sigma_e^2$. The linear relationship between Y and X can be written as $Y_i = \beta_0 + \beta_1 W_i + \varepsilon_i$,
where $\varepsilon_i = \beta_1 e_i + \epsilon_i$, $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \beta_1^2\sigma_e^2 + \sigma^2$

- **Non-linear measurement error model**

If the response or the predictor variable is measured with errors and the response variable is nonlinearly related to the explanatory variable, then the measurement error model is called a non-linear measurement model. Both the classical and Berkson non-linear measurement error models may appear separately or jointly in a single application.

2.4.3.3 Some of the methods of correcting measurement error bias

Measurement errors are cause of bias, inconsistency in parameter estimates and erroneous conclusions in statistical analyses. Consequently, many researchers have devoted considerable effort to this problem in order to design methods for addressing it. This has led to the development of a rich literature on dealing with measurement error in the response variable and covariates(e.g. Fuller (1987); Gustafson (2004); Carroll et al. (2006); Buonaccorsi (2010)). The literature on measurement error has become a cornerstone in many fields as many important risk factors are generally acknowledged to be mismeasured. Cited measurement error correction techniques include structural equation models, two stage least squares regression, partial least squares regression, ordinary least squares regression on factor scores, regression calibration, method of moments, simulation extrapolation, moment reconstruction method, disattenuated regression on summated rating scales (SRS), and multiple over-imputation.

Th structural equation modelling approach attenuates the measurement error bias by finding latent dimensions that could have been the origin of the host of mismeasured values. This method uses multiple the indicators structural equation approach to simultaneously estimate the measurement error and the parameters through the likelihood estimation. All information in the variance-covariance matrix variables is employed in the joint estimation. However, one of its drawbacks is that if there are estimation errors in some parts of the model they are likely to affect the estimation of parameters (Charles, 2005). Moreover, this method is believed to rely heavily on large samples and it might present a great degree of complexity if it is not accompanied by strong assumptions (Bisbe, Coenders, Saris, & Batista-Foguet, 2006).

Two stage least squares regression consists of two steps. In the first step, a variable which is highly correlated to the mismeasured variable but not correlated to the error term, known as an instrumental variable, is selected. In the second step, least squares regression is carried out with the instrumental variable replacing the variable measured with error in the model. Although this method alleviates the error bias, it produces estimates that are heavily dependent on the choice of instrumental variable.

Both partial least squares (PLS) regression and ordinary least squares (OLS) regression are known to be useful if the purpose of the analysis is predictive or is exploratory of summated rating scales. The partial least squares regression on SRS is consistent only under the condition of perfect reliability or the number of items per dimension tends to infinity (Bisbe et al., 2006).

As an alternative method to PLS and OLS regressions on summated rating scales, Bisbe et al. (2006) developed a disattenuated regression (DR) on SRS. This method is executed in three main steps. Firstly, the reliability of the summated rating scales is estimated. Secondly, this estimate of the reliability is used in the computation of variances of summated rating scales. Lastly, the computed variances are fed in the variance-covariance matrix from which the ordinary least squares estimates are computed.

The method of moments estimator therefore corrects the bias by simply dividing the parameter estimate by the reliability ratio. This method suffers two major drawbacks: it is hugely dependent on the estimate of the measurement error variance and it works pretty well with linear models.

The simulation extrapolation method simulates the effects of adding error to a single variable measured with error and it uses the simulated values to infer about the case of the values of an error free variable. The computation becomes difficult in case of multiple variables measured with errors. Also, it relies heavily on the extrapolated parameters.

The regression calibration method is a two stage regression method which replaces a mismeasured variable with an estimate that is a function of the variable measured with error. In other words, the calibrated data are expected values conditional on the measured data. The calibrated data is then used in the analysis instead of the observed values. This method presents a number of loopholes. It is only consistent for nonlinear regression models; it is not convenient for the estimation of residual variance analysis as the distribution of the estimated variable does not preserve the variance-covariance of the underlying unmeasured variable; and it is valid under the condition that the measurement error does not depend on the value of the response and mismeasured variable (i.e. under non-differential measurement error assumption) (Freedman, Fainberg, Kipnis, Midthune, & Carroll, 2004).

The moment reconstruction method developed by Freedman et al. (2004) is similar to regression calibration in that it replaces the observed data with adjusted values, but the adjusted values are empirical Bayes estimates of the true values conditional on the the response variable. The adjusted values have the same first two moments as the unobserved true variable data.

Blackwell, Honaker, & King (2017) developed over-imputation which is an extension of the multiple imputation method for missing data. Their method is based on a concept that views the measurement error as a type of missing data problem where the measured values are regarded as prior information for the true unobserved values. If no prior information is available, which is the extreme case of measurement error, then one has to deal with the missing data problem. The method imputes the missing values from their predictive posterior and overwrites mismeasured values or variables with draws obtained from their predictive posterior where the observed values, other variables, and available assumptions are used as prior information. This method allows one to deal with missing data problems and correct for measurement error in more than one variable.

Another way of attenuating the measurement error cited in literature is to fix the proportion of measurement error to a fixed value (Charles, 2005). Although the method may be easy to implement, it does not provide the guidelines on how the proportion of error is chosen. Instead of fixing the error to a priori value, some researchers have opted to assign a prior distribution to error component (Charles, 2005).

2.5 Disease mapping

2.5.1 Introduction

Lawson & Williams (2013) define disease mapping as a geographical distribution of a disease within a population. This distribution is achieved through the visual representation of the geographical residential addresses of diseased individuals (referenced spatial data points) or counts of individuals with a disease in small areas. A disease map, which is the collection of disease information such as residential locations of individuals or summary measures for groups of individuals in small areas, is an essential element of the disease mapping discipline as it is an efficient way of exhibiting the distribution of phenomena in space and time as well. Maps can reveal and communicate spatial findings in a better way than statistical tables do. Generally, two types of maps are distinguished, namely maps of infectious diseases and maps of non-infectious disease. The latter maps are used to point out sources or causes of outbreak referred to as putative sources of health hazard. Whilst the former type of maps are used to analyse time trends and the spatial clustering of disease and also to identify possible associations between factors and disease clusters. Disease maps are constructed either using raw data or results of some statistical analysis such as standardised mortality, morbidity ratio or relative risk (Blangiardo & Cameletti, 2015).

In this section, the focus is on the analysis of count data. The Poisson regression models are considered as standard tools for analysing count data when the key Poisson model assumption of equality between the mean and variance is met. However, count data quite often presents an excess of zero counts than it would be expected under a standard normal Poisson process and hence the equality assumption does not hold. This type of data is referred to zero inflated count data. Such data can be further classified into two classes namely, upper bounded data with excess of zeros or unbounded count data with excess of zeros. The former class is referred to as the binomial type whereas the latter class is known as a Poisson type (Hall, 2000). Data generating processes that result in zero inflated count data are commonly encountered in various fields such as agriculture, econometrics, manufacturing, road safety, medicine, sexual behaviour, and horticulture. To model the aggregated data to a large spatial area or temporal unit is seen as one of the ways of circumventing the problem posed by the excess of zeros in sparse count data. However, the aggregation of data introduces the ecological bias, which is a fallacy that occurs when conclusions about individual associations are based on aggregate (Rodriques-Motta, Gianola, Heringstad, Rosa, & Chang, 2007).

Various methods are proposed in the literature to model sparse count data dominated by zeros. In recent years, zero inflated Poisson (ZIP) and Zero inflated negative Binomial (ZINB) models have been extensively used to overcome the over-dispersion problem induced by the excess of zeros. However, according to Ridout, Demétrio, & Hinde (1998), both zero-truncated models and zero-inflate models (ZIP and ZINB) are explicitly dependent on the functional form of the probability assumed for zero counts and consequently a wrong functional form specification leads to inconsistency in parameter estimates. A test score to compare the zero-inflated regression model versus the zero-inflated negative binomial was developed by Ridout et al. (1998) in order to overcome the problem of functional form misspecification. The Rao and Chakravarti criterion to distinguish between a simple Poisson regression and ZIP models is also quite often used (Rodriques-Motta et al., 2007). Among other models proposed by scholars as remedies to excess of zeros in count data are the sparse Poisson convolution model (Song et al., 2011), standard unimodal distribution with extra dispersion, non standard mixture models, two part models (Cunningham &

Lindenmayer, 2005; Fernandes, Schmidt, & Migon, 2009) and Neyman type A distribution (Dobbie & Welsh, 2001).

2.5.2 Univariate disease mapping

Suppose $\mathbf{y}$=(y_1,. . .,y_n) is a vector of observed counts of disease cases for a set of region i=$\{1, \ldots, n\}$ in a given study domain D. In a generalised linear modelling approach, such counts are modelled as either Poisson or binomial random variables through a log or logit link function, respectively. When dealing with rare diseases, a Poisson model can be employed as an approximation to a binomial model (Gelfand et al., 2010). However, a standard Poisson model does not deal with extra-dispersion which might be caused by a spatial dependence among areal units. To take into account the spatial dependence, Poisson models which allow to borrow information across areal units are used. To this end, we outline different forms of random effects modelling which include exchangeable random effects, spatial structured random effects, and a combination of exchangeable and structured random effects. We assume that the rate of the disease is explained by random effects solely in the absence of covariates.

2.5.2.1 Exchangeable random effects modelling

Let y_i be the count of cases in the areal unit i of region D, n_i be the population count in the same areal unit i, λ be the probability of contracting the disease, $E_i = \lambda n_i$ be the corresponding expected count of cases for the areal unit i, and m_i be the relative risk for contracting the disease in the area i. Then

$$y_i|m_i \sim Poisson(E_i m_i), \tag{2.68}$$

where log (m_i)= β_0+ϕ_i, β_0 is a fixed intercept associated with the whole study domain D, and ϕ_i is a random intercept associated with each area i. The excess variability or overdispersion is introduced through the exchangeable random intercept approach (ϕ_i). That is $\phi_i \sim N(0,\sigma^2_\phi)$, for i=1,. . .,n, provided the variance is known or is assigned a proper hyperprior (e.g. inverse gamma distribution).

The introduction of the random effects allows to relate local relative risks among themselves (m_is) via the prior distribution and reduce the estimation process of n parameters to the estimation of two parameters, namely the intercept β_0, and the variance σ^2_ϕ of ϕ.

2.5.2.2 Spatially structured random effects modelling

With regards to this option, the idea is to replace the set of exchangeable priors at the second stage of hierarchical modelling with a spatially structured prior distribution. This yields local estimates which are weighted averages of area data value and observations in neighbouring areas. Thus, this modelling approach induces some form of correlation. Two approaches are distinguished: the multivariate gaussian model and the conditional autoregressive (CAR) model. In both approaches, Eq. (2.68) becomes

$$y_i|m_i \sim Poisson(E_i m_i), \tag{2.69}$$

where log (m_i)= β_0+ω_i.
The distribution of ω=(ω_1,. . . , ω_n), a vector of spatially correlated random effects is discussed differently according to the two approaches above mentioned. First, we consider that ω follows a multivariate normal distribution with mean zero and spatial covariance matrix (Σ_ω). That is $\omega \sim MVN(0, \Sigma_\omega)$. Naturally, the spatial covariance matrix comprises parametric functions defining covariance as a function of the relative locations of any pair of observations. The following are some standard families of covariance functions that meet the necessary and sufficient condition of positive definiteness:

- The Mat$\acute{e}$rn family

The Mat$\acute{e}$rn family is the most popular family that meets both criteria of a decreasing correlation between two spatial processes $S(x)$ and $S(\acute{x})$ as the euclidean distance d=$\parallel x - \acute{x} \parallel$ increases and a varying smoothness in the spatial process. It is generally favoured because of its flexibility and the physical meaning of the shape parameter (i.e. the differentiability measure of $S(x)$) (Diggle & Ribeiro Jr, 2007). The Mat$\acute{e}$rn correlation function is given by;

$$Cor_M(S(x),S(\acute{x})) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa \parallel x-\acute{x} \parallel)^{\nu}\kappa_v(\kappa \parallel x-\acute{x} \parallel), \tag{2.70}$$

where $\parallel . \parallel$ denotes the Euclidean distance, $\kappa_\nu(\cdot)$ is the modified Bessel function of second order, k and ν are the scale parameter and smoothness (shape) parameter, respectively. If ν=0.5 and $\nu \to \infty$, the Matérn correlation function becomes the exponential correlation function and the Gaussian correlation function, respectively. The mathematical expressions of the exponential and Gaussian correlation functions are given as

$$Cor(S(x),S(\acute{x})) = \exp(-\kappa \parallel x-\acute{x} \parallel) \quad (Exponential) \tag{2.71}$$

$$Cor(S(x),S(\acute{x})) = \exp\{-(\kappa \parallel x-\acute{x} \parallel)^2\} \quad (Gaussian) \tag{2.72}$$

- The powered exponential family

Like the Matérn family, the powered exponential family produces correlation functions which are monotonically declining in euclidean distance. But, it is not as flexible as the Matérn family as its Gaussian process $S(x)$ is mean-square continuous but not mean-square differentiable for all $0 < \nu < 2$. This family has a correlation function defined by

$$Cor(S(x),S(\acute{x})) = \exp\{-(\kappa \parallel x-\acute{x} \parallel)^{\nu}\}, \tag{2.73}$$

where κ and ν are scale and shape parameters respectively; and 0$\leq \nu \leq 2$.

- The spherical family

The correlation function of this class is given by

$$Cor(S(x),S(\acute{x})) = \begin{cases} 1-\frac{3}{2}\kappa \parallel x-\acute{x} \parallel +\frac{1}{2}(\kappa \parallel x-\acute{x} \parallel)^3 & \text{if } 0\leq \parallel x\text{-}\acute{x} \parallel \leq \kappa \\ 0 & \text{if } \parallel x\text{-}\acute{x} \parallel > \kappa \end{cases} \tag{2.74}$$

The spherical is commonly used in classical geostatistics. However, it is less flexible as compared to the two-parameter Matérn. One major difference of the spherical family from other families is that it has a finite range for sufficient large euclidean distance (i.e. $Cor(S(x),S(\acute{x}))$=0 for $\parallel x\text{-}\acute{x} \parallel \gg \kappa$). Also, it is only once differentiable at $\parallel x\text{-}\acute{x} \parallel = \kappa$

- Non-monotone correlation functions family

One fundamental difference between this and the other families described earlier is that it is characterised by an oscillatory behaviour. In nature, non-monotone correlation functions are scarce. One notable example of this type of correlation function is the sinusoidal function of the euclidean distance whose algebraic form is given by

$$Cor(S(x), S(\acute{x})) = (\kappa \parallel x - \acute{x} \parallel)^{-1} \sin((\kappa \parallel x - \acute{x} \parallel) \tag{2.75}$$

Second, unlike in the preceding section wherein the specification of the variance-covariance matrix was achieved through a direct use of parametric functions of distance; in this section its specification is accomplished by using spatial proximity measures. The structure considered here emulates the time-series approach which commonly uses autoregressive models where the current observation is regressed on observed values of a subset of observations that have occurred in the recent past. In the spatial context, the observations that have occurred in the recent past are equivalent to observations that have occurred nearby. In simple terms, an autoregressive model reflects a self-regression model. Through such regression, the spatial similarity is introduced by treating observations at neighbouring locations as additional covariates in the model, instead of formulating an explicit mathematical expression for the covariance function of the error terms. Two classes of autoregressive models, namely the conditional autoregressive (CAR) model and the simultaneous autoregressive (SAR) model, are reviewed below.

- Conditional autoregressive (CAR) model:

Assume that ω_i, the area-specific effect specified in Eq. (2.69) is a normally distributed random variable as follows

$$\omega_i \mid \omega_{-i} \sim N(\mu_i + \frac{\rho}{N_i} \sum_{j=1}^{n} b_{ij}(\omega_j - \mu_j), \sigma_i^2) \tag{2.76}$$

$\sigma_i^2 = \frac{\sigma_u^2}{N_i}$ where ω_{-i} denotes the vector of all area-specific effects in the neighbourhood of area i except ω_i, N_i is the number of the neighbours of area i,

μ_i is the mean for area i which is a weighted average of the other ω_j (for $j \neq i$) and $\sigma_i^2 = \frac{\sigma_u^2}{N_i}$ is its variance, b_{ij} defines the neighbourhood spatial proximity (b_{ij}=1 if the area i is neighbour of area j and 0 otherwise, b_{ii}=0), and ρ is the parameter that controls the properness of the distribution. The Eq. (2.76) is referred to as a proper conditional autoregressive model of $\omega_i \mid \omega_{-i}$. The joint proper conditional autoregressive model for ω=(ω_1,. . .,ω_n) is given by

$$\omega \sim N(\mu, (I - \rho B)^{-1}\sigma^2), \tag{2.77}$$

where μ=(μ_1,. . . ,μ_n), B is the matrix generated by the elements $\frac{b_{ij}}{N_i}$, and σ=diag(σ_1,. . .,σ_n). The nonsingularity of the matrix $(I - \rho B)^{-1}\sigma^2$ is assured if $\rho \in (\frac{1}{\lambda_{(1)}}, \ldots, \frac{1}{\lambda_{(n)}})$(Banerjee et al., 2004), where $\frac{1}{\lambda_{(i)}}$s are the eigenvalues of B. If ρ=1 in Eq. (2.76), then the proper CAR model simplidies to a version known as an intrinsic autoregressive (iCAR) model represented by

$$\omega_i \mid \omega_{-i} \sim N(\mu_i + \frac{1}{N_i}\sum_{j=1}^{n} b_{ij}(\omega_j - \mu_j), \sigma_i^2) \tag{2.78}$$

It is impossible to construct a joint distribution of a vector of ω_i that follows iCAR as the covariance matrix $(I - \rho B)^{-1}\sigma^2$ becomes singular (not positive definite).

- Simultaneous autoregressive (SAR) model:

A typical simulataneous autoregressive model for variable ω_i presented in Eq.(2.69) is expressed as

$$\omega_i = \mu_i + \sum_{j=1}^{n} b_{ij}(\omega_j - \mu_j) + \upsilon_i, i = 1, \ldots, n, \tag{2.79}$$

where the vector of residual errors υ =(υ_1,. . .,υ_n)$\sim$ $N(0, \Sigma_\upsilon)$ with
Σ_υ=diag(σ_1,. . .,σ_n). From the literature the Eq. (2.79) is called the simultaneous autoregressive model, where the word simultaneous refers to the simultaneous application of the equation to each ω_i of area i. Further details on SAR and CAR models and their similarities can be found elsewhere (e.g. Waller & Gotway (2004); Banerjee et al. (2004)).

2.5.2.3 Convolution priors

When both global and local borrowing of information are included within the same Poisson model through the introduction of both exchangeable and CAR random effects for each area, the specification originates the Besag-York-Molliѐ (BYM) model expressed as follows

$$y_i|m_i \sim Poisson(E_i m_i), \quad log(m_i) = \beta_0 + \phi_i + \omega_i \tag{2.80}$$

where $\phi_i \sim N(0,\sigma_\phi^2)$ and $\omega_i \mid \omega_{-i} \sim \mathrm{N}(\mu_i + \frac{\rho}{N_i} \sum_{j=1}^{n} b_{ij}(\omega_j\text{-}\mu_j),\sigma_i^2)$, respectively. Over decades, it has become the most popular choice in disease mapping especially when estimating relative risks in small areas or adjusting for covariates effects. Although BYM is quite flexible, its related problem is that the structured and unstructured components are not easily identifiable from each other. For identifiability purposes, some care is required in assigning hyperprior distributions to the conditional variance (σ_i^2) and the marginal variance (σ_ϕ^2). For instance, if noninformative hyperpriors are assigned to both hyperparameters, then only the sum of the random effects (ϕ_i+ω_i), and not their individual values, will be identified. As a rule of thumb, it is suggested to choose the prior marginal standard deviation of ϕ_i to be approximately equal to 1.4 fold the conditional standard deviation of ω_i (Gelfand et al., 2010). Other alternative model formulations to overcome identifiability between unstructured and structured random effects include Leroux and Dean models (Riebler, Srbye, Simpson, & Rue, 2016).

2.5.3 Multiple diseases mapping

Thus far, we have reviewed some modelling methods applicable to areal counts of a single disease. However, it is quite often possible to encounter situations whereby counts of multiple diseases are observed over the same study domain. Diseases observed over the same regions may relate in different ways. For instance, they may have some common risk factors or the presence of one disease is a precursor of another disease or it may obstruct the presence of another. With univariate models, it is impossible to explore correlation structures across diseases. Thus, a multivariate spatial modelling approach will be appropriate when multiple diseases are present over a study region as it does not only allow modelling of dependence among those diseases, but also it maintains spatial dependence between areal units. To this end, various ways have been used to model jointly multiple diseases. Some of the techniques found in the literature include the separable modelling based on specification cross-covariance functions (e.g. spatial regression models, cokriging); coregionalisation modelling (e.g. intrinsic specification); shared component modelling; and multivariate Conditional autoregressive (MCAR) modelling approach (Banerjee et al., 2004). In subsequent paragraphs, we briefly review multivariate CAR and shared component modelling approaches.

2.5.3.1 Multivariate conditional autoregressive (MCAR) modelling

For decades, conditional autoregressive modelling specifications have been extensively applied to analyse mostly univariate cases of spatial data. Gelfand & Vounatsou (2003) generalised the univariate CAR to multivariate conditional autoregressive models. The generalisation was achieved by introducing spatial autoregression parameters to ensure the distributional propriety under a separability assumption, which allows to model between diseases correlations as well as the spatial dependence across space.

Let y_{ik} be the count of cases of disease k in the areal unit i of region D, $i = 1, \ldots, I$, $j = 1, \ldots, p$, E_{ik} be the expected count of cases for disease k in the areal unit i, and m_{ik} be the relative risk for contracting the disease k in the area i. The first level of the hierarchical model is expressed as follows.

$$y_{ik}|m_{ik} \sim Poisson(E_{ik}m_{ik}), \tag{2.81}$$

where log $(m_{ik})= \beta_j+\phi_{ik}$ (assuming no covariates in the model), β_j is a disease specific parameter coefficient, and ϕ_{ik} is a random effect associated with disease k in the area i. Under the separability assumption, the association structure separates into a nonspatial and spatial component (Gelfand et al., 2010). That is, the joint distribution of ϕ is assumed to be

$$\phi \sim N_{np}(0, [\Lambda \otimes (D - \alpha W)]^{-1}), \tag{2.82}$$

where $\phi = (\acute{\phi}_1, \ldots, \acute{\phi}_p\acute{)}$, $\phi_j = (\acute{\phi}_{1j}, \ldots, \acute{\phi}_{Ij}\acute{)}$, Λ is a $p \times p$ positive definite matrix, which represents the nonspatial precision matrix between p diseases (i.e. inverse dispersion), $\otimes$ denotes the chronecker product, W is a proximity matrix with elements w_{ij} (measure the closeness of areas i and j), D is a diagonal matrix with i^{th} diagonal element equal to $\sum_j w_{ij}$, and $\alpha \in [0, 1]$ is a spatial autocorrelation parameter, which ensures the propriety of the joint distribution. The distribution in Eq. (2.82) is denoted as $MCAR(\alpha, \Lambda)$. If $\alpha = 1$, the distribution simplifies to improper MCAR denoted by $MCAR(1, \Lambda)$ and referred to multivarite intrinsic autoregressive model($MIAR$).

By letting $R_j \acute{R}_j = D - \alpha W$, $j = 1, \ldots, p$, the $MCAR(\alpha, \Lambda)$ can be generalised to accommodate different smoothing parameters for each disease. The model becomes $MCAR(\alpha_1, \ldots, \alpha_p, \Lambda)$, which is expressed as

$$\phi \sim N_{np}(0, \Big[Diag(R_1, \ldots, R_p)(\Lambda \otimes I_{n \times n})Diag(R_1, \ldots, R_p\acute{)}\Big]^{-1}) \tag{2.83}$$

2.5.3.2 Shared component modelling

For areal data, multivariate models are commonly employed with the main purpose of introducing multiple dependent spatial random effects associated with areal units. Apart from the MCAR modelling approach reviewed in the above section, other approaches include a twofold CAR model, shared component, and MCMC blocking approaches which jointly model three sets of spatial random effects through three independent CAR prior distributions in a shared component model setting. In this section, we briefly reviewed the shared component approach.

The shared component model was pioneered by Knorr-Held & Best (2001). It splits the disease profile into two components, namely the disease-specific component representing spatially varying factors, and the shared component which is a proxy of unobserved spatially varying factors that are common to both or all diseases (Knorr-Held & Best, 2001). In situations where two or more diseases are observed on the same areas, modelling them jointly is a better alternative way of disease mapping instead of fitting a separate model for each disease.

The model formulation given below follows the formulation of shared component as presented by Knorr-Held & Best (2001). Let y_{ik} be the count of cases of disease k ($k = 1, 2$) in the areal unit i of region D, $i = 1, \ldots, I$ and E_{ik} be the expected count of cases for disease k in the areal unit i, and m_{ik} be the relative risk for contracting the disease k in the area i. The first level of the hierarchical models were expressed as follows

$$y_{i1}|m_{i1} \sim Poisson(E_{i1}exp(m_{i1})) \quad and \quad y_{i2}|m_{i2} \sim Poisson(E_{i2}exp(m_{i2})) \qquad (2.84)$$

In the second stage, the log of relative risks are modelled as
$m_{i1}= \lambda_i^{\delta}+\phi_{i1}$ and $m_{i2}= \lambda_i^{\frac{1}{\delta}}+\phi_{i2}$.

The three components λ_i^{δ}, ϕ_{i1}, and ϕ_{i2} were assumed to be independent. λ_i^{δ} is a shared component and its contribution to overall relative risk is weighted by δ (scaling parameter), and the other two components are disease specific components. In this model, the mean of λ was assumed to follow a flat prior, whereas for the other two components of the means were set to zero for identifiability purposes. The variances of the three components were assigned inverse gamma priors. The scale parameter, which allows a different gradient to be associated with the shared component for each disease, assumed a prior with a zero mean and variance σ_{δ}^2.

Ngesa (2014) adopted the shared component models to suit a Bernoulli process and the following models were formulated.
$log(p_{ij1})= \beta_1+\lambda_i\delta+\phi_{i1}$ and $log(p_{ij2})= \beta_1+ \frac{\lambda_i}{\delta}{}_i$,
where p_{ijk} is the probability of individual j in area i to get disease k.

In the subsequent paragraphs, the shared component models were adopted to suit data for this study through the stochastic partial differential equations (SPDE) modelling approach. Considering the bivariate model, which pools the two datasets, let y_{ij} be a binary indicator of HIV incidence at location i from dataset ($j = 1, 2$). Then y_{ij} follows a $Bernoulli(p_{ij})$, p_{ij} is the probability of a recorded HIV incident pertaining to the j^{th} dataset.

The bivariate model is then given by

$$logit(p_{i1}) = \beta_0 + \sum_{k}^{r} \beta_k X_{i1k} + f_1(g_{i1}) + z_1(s_i), \tag{2.85}$$

$$logit(p_{i2}) = \beta_0 + \sum_{k}^{r} \beta_k X_{i2k} + f_1(g_{i2}) + z_2(s_i) + \gamma z_1(s_i), \tag{2.86}$$

with $j = 1$ for NHSS and $j = 2$ for DHS where X is the vector of linear covariates with corresponding regression parameters β;
g_{ij} is the vector of ages which are assumed to follow a random walk of order 1; $z_1(s_i)$ is a Gaussian random field shared between both responses, and the interaction parameter γ describes how much of the structure captured in $z_1(s_i)$ is also inherent to the $logit(p_{i2})$.

2.6 Conclusion

In summary, this chapter reviewed the basics of Bayesian modelling that include different prior distributions and estimation methods. Also, it provided a review of spatial and spatio-temporal modelling approaches of lattice and geostatistical data. However, no review on spatial and spatio-temporal modelling techniques of point patterns was considered in this dissertation. Furthermore, some current issues in spatial and spatio-temporal modelling, which include the misalignment problem, edge effects, and measurement error models were reviewed. Lastly, univariate and multivariate disease mapping were reviewed with an inclination on only mapping of lattice and geostatistical data.

Chapter 3

Modelling spatial patterns of misaligned disease data: An application on measles incidence in Namibia

[1]

Introduction

Quite often disease data are available in aggregated formats mostly to maintain confidentiality. This leads to a misalignment problem when the goal is to analyse risk at a different level of spatial resolution different from the original administrative level where data were available.

Objective: To estimate and map the risk of measles at a sub-regional level in Namibia using data obtained at a regional level.

Methods

Using measles data from Namibia for the period 2005-2014, both multi-step and direct approaches were applied to correct for misalignment. Subsequently, ecological Bayesian regression models were fit and compared.

Results

Results showed that the variables standardised birth rate, counts of measles cases for previous year, unemployment rate and proportion of vaccinated children against

[1] Published as: (Ntirampeba, D., Neema, I., & Kazembe, L. (2017)

measles by age 12 months were significant determinants of measles risk. Constituencies having elevated measles risk were identified mostly in the northern corridor with Angola.

Conclusion

We recommend that relevant authorities should make geographical target intervention and redesign prevention and control strategies based on these findings.

3.1 Introduction

Measles is a disease caused by a highly contagious human pathogen that belongs to the Paramyxoviridae family (Bhella et al., 2007). The disease spreads through coughing, sneezing, near contact or direct contact with infected nasal or throat secretions. It has an incubation period located between 9 and 12 days and an infectivity period located between 4 and 9 days (Doungmo, Oukouomi, & Mugisha, 2014). Deaths due to measles are quite common among malnourished children and people whose immune system has been weakened by diseases that include human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS). Measles leads to other complications such as blindness, brain swelling (encephalitis), diarrhoea, ear infections and respiratory infection such as pneumonia. High death rates are commonly registered in developing countries with low per capita income and poor health service systems (WHO, 2014).

Worldwide, measles is ranked among the leading causes of mortality especially among children in developing countries. For instance, in 2013, about 145,700 deaths were recorded (WHO, 2015). Until now, there is no antiviral treatment for the measles virus. Thus far, measles vaccination and supportive care that includes good nutrition and adequate fluid intake have been used to fight measles (WHO, 2015). However, a reduction of global funding by the governments and partners has largely affected the immunisation campaigns, which has hampered efforts for a complete elimination of measles WHO (2014). Consequently, measles cases are still reported in many countries, with Angola, Ethiopia, Namibia, Bosnia and Herzegovina, Georgia, Sri Lanka and Philippines ranked among the top ten countries with high annualised measles incidence per 100,000 inhabitants in 2014 (WHO, 2017).

In Namibia, as in many countries, diseases surveillance data are often analysed in the form of aggregated data at health district or regional level because of confidentiality issues. However, health decisions might be needed at lower political boundaries such as constituencies. Nevertheless, direct inference at such lower level made on the basis of regionally aggregated data may lead to the spatial misalignment problem (Banerjee et al., 2004).

In brief, spatial misalignment appears through various processes. The first process is when the purpose of the analysis is to make inference about new points based on available information at different points or locations. This is known as point-to-point change of support. The second is when a researcher might be interested at predicting values at block level using information available at point level. This is called the point-to-block change of support. In the third process, one might seek to make inference from block values to point level, and this is referred to as the block to point change of support (Banerjee et al., 2004). In this scenario, it is inappropriate to infer about the relationships between variables at individual level using information observed at area level, as the accuracy at area and point levels is not a one-to-one relationship. This challenge is referred to as ecological fallacy. Fourth, spatial misalignment can arise when the purpose of the spatial analysis is the interpolation at new aggregation level that is different from a level where data were observed. Scholars refer to this as the modifiable area unit problem.

Various methods for resolving misalignment have been proposed (Goovaerts, 2008; Keil et al., 2013; Sturrock et al., 2014; Araújo et al., 2005; Lee & Sarran, 2015; Finley et al., 2014; Illian et al., 2009). For instance, methods have been applied to downscale the distribution of data from coarse to fine grain, and that include direct method, point sampling method and hierarchical Bayesian method. These methods have been adopted to deal with this scenario of misalignment (Keil et al., 2013; Sturrock et al., 2014; Araújo et al., 2005). Other techniques have been developed to deal with spatial misalignment that arises when the response variable is available at bigger irregular shaped area units and covariates are available at smaller fine grids (Lee & Sarran, 2015), in which a multi-step approach has been applied. In the case where misalignment occurs with non-nested overlapping grids, hierarchical Bayesian approaches have been employed (Banerjee et al., 2004; Finley et al., 2014). Recently, the latter has been extensively applied as it permits to derive posterior predictive distributions for both parameters and the epidemiological outcome of interest. It is also suitable when dealing with multiple sources of uncertainty and it enables to incorporate additional sources of information in the form of prior knowledge (Illian et al., 2009).

The aim of this study was two-fold. First, the study aimed to identify constituencies (sub-regions) in Namibia that have an elevated risk of measles and also to visualize smoothed patterns of risk of measles. Second, the study aimed to determine factors associated with the risk of measles in Namibia.

3.2 Methods

3.2.1 Sources of data

Measles cases were abstracted from the health management information system (HMIS) database within the Ministry of Health and Social Services (MoHSS) in Namibia. The database included all suspected measles cases from which positive cases were extracted. Any patient consulting a health facility becomes a suspected case if the patient is diagnosed with fever and generalised maculopular rash lasting for three days or longer, and a cough, coryza or conjunctivitis. Such a case will be investigated and adequate blood specimen is collected and examined at the Namibia Institute of Pathology (NIP). If the blood specimen is found to have serological confirmation of a recent virus infection, the case is classified as laboratory confirmed. However, there are other cases wherein blood specimens are not taken for serological confirmation, but they are linked to laboratory confirmed cases. Such cases are known as epidemiologically confirmed. A suspected case is discarded if it has been completely investigated or the blood specimen is declared by NIP as not having serological evidence of recent measles virus. This determination of a measles case is based on the World Health Organisation's (WHO) standard definition, which considers a measles case as either an epidemiologically confirmed case or a laboratory confirmed case (WHO, 2015; Heymann, 2015). Although the HMIS database has information from 2001 to partly 2015, the 2005−2014 period provided consistent information for the entire country, and hence only data from this period were considered in this study.

Covariates used in this study were obtained from the 2011 Namibia population and housing census (NPHC) and the 2013 Namibia demographic health survey (NDHS). The variables included proxies of social mixing patterns (average household size and proportions of children attending pre-primary and schools), unemployment rates and birth rates. Table 3.1 provides a list of all variables used in the analysis, as identified through literature (Doungmo, Oukouomi, & Mugisha, 2014; Zagheni et al., 2008; Held, Höhle, & Hofmann, 2005; Adika, Baralate, Agada, & Nneoma, 2013; Jasem, Marof, Nawar, & Islam, 2012; Mayet et al., 2013; Beyene, Tegegne, Wayessa, & Enqueselassie, 2016). Shapefiles that defined the administrative boundary maps were also obtained from the Namibia Statistics Agency (NSA). Although the administrative boundaries have changed over time, in this study we have used the 2011 administrative boundaries that match with variables derived from the 2011 NPHC.

Table 3.1: Description of variables considered for the analysis

Variable	**Variable name**
1	Standardised average household size
2	Counts of measles for previous year (2004)
3	Unemployment rates
4	Standardised birth rates
5	Proportions of children attending schools
6	Proportion of vaccinated children against measles by age 12 months
7	Proportions of children attending pre-primary

3.2.2 Statistical methods

Each of the 13 regions in Namibia is sub-divided into constituencies, giving a total of 107 sub-regions. The counts of measles cases are available at 13 regions, of which our aim was to estimate the risk at constituency level. This introduces the problem of misalignment in the analysis. To overcome misalignment, two approaches (multi-step and direct methods) are used.

3.2.2.1 Multi-step approach

This method allows the allocation of region/district disease totals to constituency proportional to the constituency area or population. However, the areal proportional allocation method assumes that the population is uniformly distributed throughout the entire area. Namibia is a semi-desert country, and its population is not spread uniformly throughout the territory; rather, people are living in towns or settlements. It has been shown that measles infection is proportional to the size of the population in each location (Doungmo et al., 2014). Thus, the population proportional allocation was applied. Steps of the multi-step method are as follows:
(i) Overlay constituencies on regions. This enables to determine exactly what proportion of a given constituency is susceptible or infected by measles.

(ii) Find all total values of measles cases for all constituencies. The computation of these values is based on the population proportional allocation concept, as it has been shown to be more appropriate relative to areal proportional allocation for infectious diseases (Doungmo et al., 2014). This is formulated as follows: $y_{ik} = \frac{p_{ik}}{P_k} Y_K$, where y_{ik} is the number of measles cases in the constituency i of region k; Y_K is the number of measles cases in the k^{th} region that contains the constituency i; p_{ik} is the total population of constituency i included in the region k; and P_k is the total population of the region k.

(iii) Apply spatial smoothing techniques to the computed measles cases. To this end, we explored the Poisson and negative binomial hierarchical regression modelling approach. These models are ideal for count data, and by applying hierarchical models, we incorporate covariate information and any other sources of uncertainty in the parameter estimation process.

The model formulation given below follows the formulation of nested block-level modelling presented by Banerjee et al. (2004). In brief, let I be the number of constituencies overlaid in a given region and K be the number of regions in Namibia, y_{ik} the total measles count in the constituency i of region k, such that $i = 1, \ldots, I$ and $K = 1, \ldots, K$, n_{ik} is the population count in constituency i of the k^{th} region, λ is the probability of contracting the disease. In this study, λ is assumed to be the 2014 Namibia annualised measles incidence per 100,000 inhabitants such that $E_{ik} = \lambda n_{ik}$ is the corresponding expected disease count for constituency i , m_{ik} is the relative risk for contracting the disease in the constituency i , and X_{ik} are covariates present in constituency i.

The first stage model for disease counts is given by a Poisson such that

$$y_{ik}|m_{ik} \sim Poisson(E_{ik}m_{ik}), \tag{3.1}$$

where

$$log(m_{ik}) = log(E_{ik}) + X_{ik}^T\beta + \phi_{ik} + w_{ik}, \tag{3.2}$$

where in Eq. (3.2) during the second stage of Bayesian hierarchical modelling, we specify the distribution of m_{ik} as a function of the covariates (X_{ik}^T) in the constituency i for some fixed effects, β, and spatial random effects, ϕ and w_{ik}. Alternatively, a negative binomial (NB) can be specified as

$$y_{ik} \sim NB(n_{ik}, p) \tag{3.3}$$

where

$$p = 1 - exp[X_{ik}^T\beta + \phi_{ik} + w_{ik}] \tag{3.4}$$

for $y_{ik} = 0, 1, \ldots$ and n_{ik} is the number of individuals at risk with a probability of getting measles p. Similarly, in Eq.(3.4), p is modelled as a function of covariates and some spatial random effects in similar a way as in Eq. (3.2).

Estimation of the models (3.2) and (3.4) follows a Bayesian inference approach. As such, prior assumptions need to be specified. For fixed effect parameters β , weak informative Gaussian priors $\beta \sim N(0, \tau^{-1}I)$ with small precision τ on identity matrix are assumed. The term, ϕ_{ik} assumes a prior normal $N(0, \frac{1}{\tau_h})$ and it controls the global extra-Poisson variability in relative risks or captures constituency-wide heterogeneity. This prior is also referred to as an independent and identically distributed (IID) prior, such that the effect ϕ_{ik} for each constituency is independent of all other constituencies.

On the other hand, w_{ik} captures the extra-Poisson variation in relative risks that vary locally (spatial clustering) and w_{ik} are assumed to be distributed according to the intrinsic conditional autoregressive (ICAR) model. We specify the ICAR model as follows

$$w_{ik}|w_{jk,i \neq j} \sim N(\frac{1}{n_i}\sum_{ik} w_{ik}, \frac{1}{\tau_{n_i}}), \tag{3.5}$$

where τ_{n_i} and n_i are the precision parameter and the number of neighbours of constituency j respectively (Besag et al., 1991). Under this prior, the effect of w_{ik} for each constituency is normally distributed with a mean effect equalling the average effect of effects of neighbours of constituency i with τ_{n_i} precision. The neighbours are defined in terms of constituencies sharing at least one point (queen adjacency). τ,τ_h, and τ_{n_i} assume inverse gamma prior distributions.

3.2.2.2 Direct approach

This is one of the models that are commonly used in ecology to downscale the distribution of species from coarse scale to fine scale (Keil et al., 2013; Sturrock et al., 2014; Araújo et al., 2005). The conventionally used direct approach (Keil et al., 2013; Sturrock et al., 2014; Araújo et al., 2005) assumes that the cases distribution at the fine scale (constituency level) is driven by the same processes as at the coarse scale (regional/district level). Thus, the method fits a hierarchical spatial regression model at coarse scale and then the estimated parameters are used in the spatial regression at fine scale.

Specifically, in this study, we fit the following model

$$y_i|m_i \sim Poisson(E_i m_i) \tag{3.6}$$

$$log(m_i) = log(E_i) + X_i^T \beta + \phi_i + w_i \tag{3.7}$$

where in model (3.7) the distribution of the relative risk m_i was specified as a function of the covariates X_i^T (mean values of covariates at constituency level) in the region i for some fixed effects, β, and some spatial random effects, ϕ_i (unstructured spatial random effect), and w_i (structured random effect). The prior and hyper-prior distributions were specified in the same way as in the aforementioned multi-step approach. The fixed effects β are then used directly in Eq. (3.8) in order to predict the specific constituency relative risk m_{ik} through the following equation

$$log(m_{ik}) = X_{ik}^T \beta \tag{3.8}$$

To predict the number of cases in a given constituency (fine scale level), values obtained from model (3.8) will be fed in the following model:

$$y_{ik}|m_{ik} \sim Poisson(E_{ik} m_{ik}) \tag{3.9}$$

3.3 Results

3.3.1 Exploratory analyis

Before fitting the models to the data, we explored the issues of multicollinearity and spatial autocorrelation that may arise in the data. Firstly, a multicollinearity is a condition that happens when independent variables are highly correlated. This condition affects the estimated regression coefficients of independent variables, as their sampling errors tend to be large. Many scholars still insist that there is no clear critical value of correlation among independent variables to signal multicollinearity (Keller, 2012). A common way to measure multicollinearity is to use the variance inflation factor. Generally, the multicollinearity between independent variables is regarded to be severe if the largest variance inflation factor is greater than 10.

To avoid multicollinearity, we ran a correlation analysis for the variables (Table 3.1) and computed the variance inflation factor. It was found that the variables standardised average household size and proportions of children attending schools ($r = 0.652$), proportions of children attending schools and proportions of children attending pre-primary ($r = 0.6452$), and standardised average household size and proportions of children attending pre-primary ($r = 0.552$) were highly correlated. The variance inflation factor was 11.5. Consequently, only the variable standardised average household size as a proxy of social mixing was used. We then fitted 12 models, which are summarized in Table 3.2. The first three models are for both Poisson and negative binomial distributions that assume the variability in measles incidences is solely due to spatial random effects.

Secondly, we used the global Moran's I statistic determine the overall strength of spatial dependence. This statistic is a useful measure of the overall clustering. The global Moran's statistic was 0.157 (p-value=0.0048) with a variance of 0.00415. Thus, this positive significant Moran's I value implies that values in neighboring constituencies tend to cluster. To detect local spatial patterns, we have used local Moran's I. It enhances to identify clusters, which are observations with very similar neighbours and hotspots, which are characterised by observations with very different neighbours. Fig. 3.1 (a) shows the map of local Moran's I statistics. From this figure, it can be noted that constituencies in Opuwo have elevated positive Moran's I values. Fig. 3.1 (b) shows the probability values associated with local Moran's I statistics. It indicates the constituencies in Opuwo have significant local Moran's I values. These results have inspired the spatial analysis undertaken in the subsequent sections.

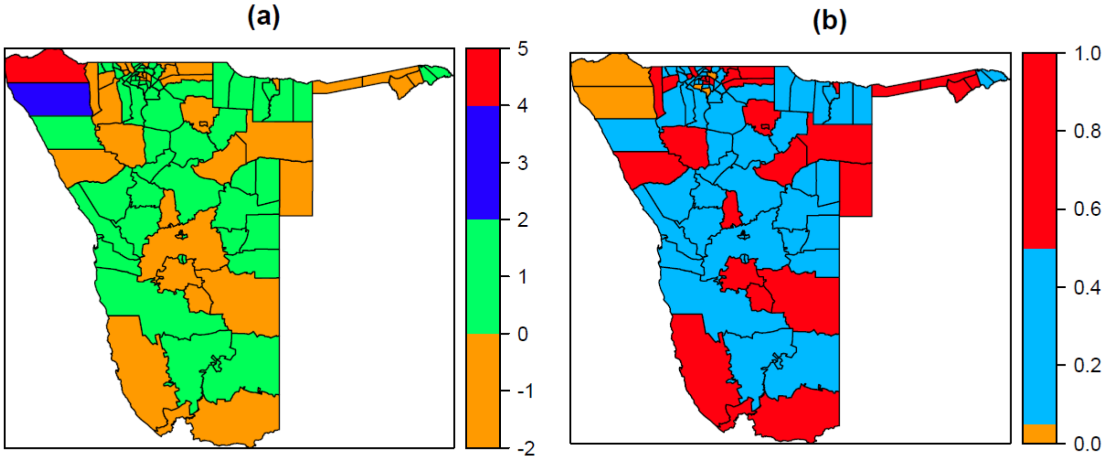

Figure 3.1: Distributions of (a)local Moran's I (observed) values and (b)probability values of Moran's I for all constituencies

3.3.2 Model selection

Model 1 explores only IID random effects. Model 2 assumes for each region a spatially structured random effect through ICAR model known as the Besag model. The third model (Model 3) is the convolution model that assumes for each region two components of random effect, namely, unstructured IID and ICAR; in other words, this is a Besag-York-Mollie (BYM) model. The other three models for both distributions were obtained by adding covariates to models 1-3 in order to assess effects of covariates on the risk of measles. The best model was identified using the deviance information criterion (DIC). The DIC is given by $DIC = D + 2p$, where D the deviance is evaluated at the posterior mean and p is the effective number of parameters. By the rule of thumb, the best model is one with the smallest DIC.

The significance of parameters was assessed using credible intervals. Generally, if a credible interval for θ does not contain zero, then the parameter is statistically significant. In Bayesian setting, a $100(1-\alpha)\%$ credible interval for θ is an interval (a, b) such that $(a \leq b \mid o_1, \ldots, o_n) > (1-\alpha)100\%$, where α is a small value between 0 and 1, and $o_1, \ldots, o_n$ are observed sample values. It is the analog of confidence interval in the classical approach. When data have been observed, the credible interval is fixed, while θ is random. This is in dissimilarity to the classical confidence interval where the interval is random while θ is a fixed parameter. The interpretation of a credible interval is different from the one of the classical interval. In the Bayesian paradigm, the credible interval is interpreted as "the probability is at least $(1-\alpha)100\%$ that θ lies within the interval(a, b)". In classical approach, the confidence interval is interpreted as "$(1-\alpha)100\%$ of all such intervals(a, b) will contain the true parameter θ". To estimate Bayesian posterior marginal distributions and any other posterior inferences for all the 12 models, the integrated nested Laplace approximation (INLA) approach was used. Model fitting was carried out in R statistical software (R Core Team, 2017).

Table 3.2 shows that Models 2 and 3 (Poisson models), which took into account the random spatial variation, are the best competing models ($DICs$: 849.95 and 850.07, respectively) among those that did not include covariates. This reveals the presence of spatial clustering or correlation in measles risk. After adding covariates in models, it was noted that Models 5 and 6 were equally good fitting for these data. Model 5 was used to generate the relative risks (RR) for contracting the measles virus and the probabilities to assess constituencies with elevated relative risks.

Table 3.2: Summary of models fitted to measles data for Namibia and their corresponding DICs

Poisson models					
Model	spatial component	Fixed	D	p_d	DIC
1	IID	-	657.58	99.43	856.44
2	ICAR	-	661.07	99.44	849.95
3	CAR	-	660.93	94.57	850.07
4	IID	All covariates	658.15	97.93	854.01
5	ICAR	All covariates	662.3	93.27	848.84
6	CAR	All covariates	662.4	93.29	848.94
Negative binomial models					
1	IID	-	1081.72	1.96	1085.64
2	ICAR	-	939.14	43.15	1025.44
3	CAR	-	938.78	43.42	1025.62
4	IID	All covariates	1060.39	6.93	1074.25
5	ICAR	All covariates	939.47	42.06	1023.59
6	CAR	All covariates	941.41	41.16	1023.73

3.3.3 Fixed effects

Table 3.3 presents a summary of fixed effects for all variables included in the model. In summary, from 95% credible intervals, we observed that the standardised birth rates, counts of measles for previous year (2004) and unemployment rates had significant positive effects on measles incidence, whereas the proportion of vaccinated children against measles by age 12 months had significant negative association with risk of measles. However, the standardised average household size did not show any significant association, although this was positively related to the risk of measles.

Table 3.3: Summary of models fitted to measles data for Namibia and their corresponding DICs

Fixed effects	Posterior mean	Standard deviation	95%
Standardised average household size	0.0523	0.063	(-0.072,0.176)
Counts of measles for previous year (2004)	0.017	0.006	(0.0043,0.0294)
Unemployment rates	0.007	0.003	(0.0011,0.0129)
Standardised birth rates	0.143	0.048	(0.0494,0.2364)
Proportion of vaccinated children against measles by age 12 months	-0.005	0.003	(-0.0099,-0.0001)

3.3.4 Spatial distribution of measles relative risks

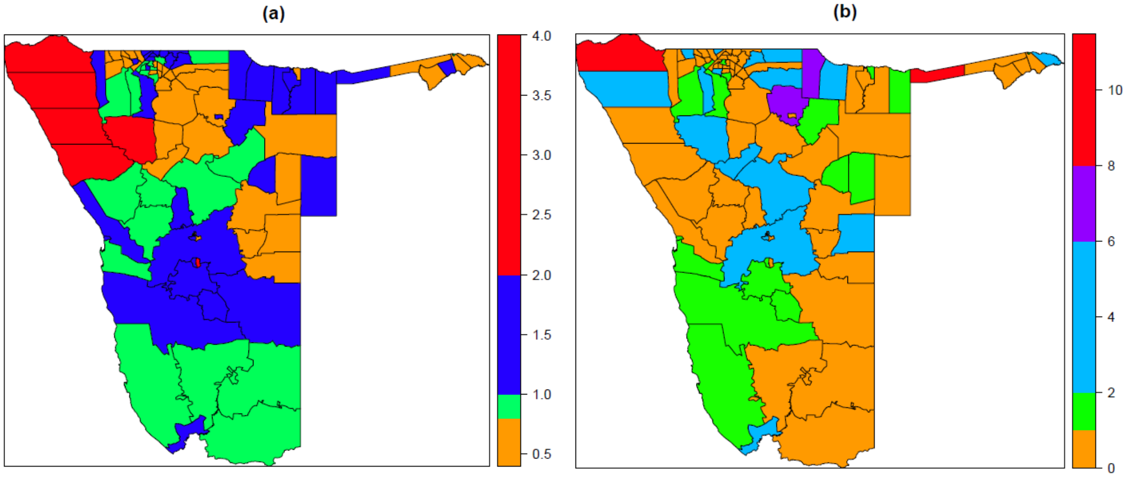

Figure 3.2: Distributions of constituency specific relative risks obtained from: (a)Multi-step approach and (b)Direct approach

In disease mapping, the most important aspect is to determine the areas with excess risks. Maps (a) and (b) in Fig. 3.2 show the distributions of constituency specific relative risks obtained from the multi-step approach and direct approach, respectively. Map (a) (Fig. 3.2) indicated that the Kunene region has constituencies with high residual relative risks, whereas map (b)(Fig. 3.2) showed that Epupa (Kunene), Mungu and Mukwe (Kavango), Guinas and Tsumeb (Oshikoto) and Omatako and Okahandja (Otjzondjupa) constituencies had high relatives.

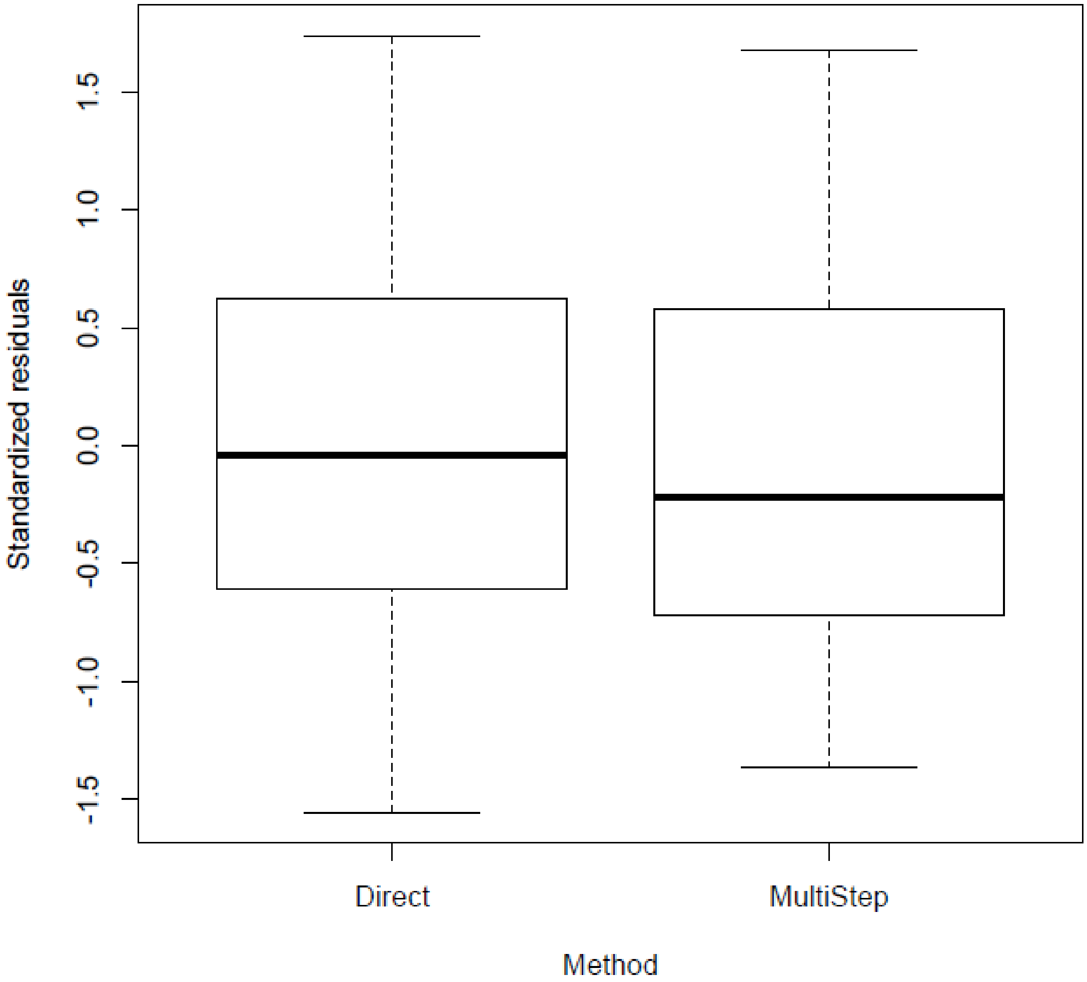

Figure 3.3: Boxplot of standardised residuals obtained from the direct and multi-step approach models

3.3.5 Modelling approach comparison

While the multi-step modelling approach provides a DIC statistic, the direct method does not provide such statistic. Thus, the comparison of the models resulting from these modelling approaches cannot be achieved through a DIC statistic. The analysis of residuals has been used as an exploratory tool to assess which model performs better. Fig. 3.3 gives a boxplot of standardised residuals obtained from the direct and multi-step approach models, respectively. This figure indicates that the multi-step approach model provides lower median standardised residuals. In addition, the models were compared using the root mean square error (RMSE) statistic. It was found that the RMSE associated with the multi-step approach model was small relative to one of the direct models (i.e. 6.40 versus 6.60), thus confirming the boxplot results that the multi-step approach model provided a relatively better model.

3.4 Discussion

The main aim of this study was to use aggregated data obtained at regional level to estimate and map the risk of measles at a lower level (constituency level). To achieve this, we corrected for spatial misalignment using both direct and multi-step approach methods. Subsequently, a spatial Poisson regression model was applied to explain the variation of measles risk in Namibia. The model thus developed included socio-economic covariates that explained the risks of measles in Namibia.

Findings showed that the measles risk varied remarkably (3.2). Using either the direct or multi-step approaches, constituencies of high risk were observed along the borders with Angola, notably in Kunene region (i.e. Opuwo, Sesfontein, Khorixas, Kamanjab and Outjo constituencies) and Kavango region (i.e. Epupa, Mukwe and Mpungu constituencies). This could be due to the free movement of people to and from Angola, whereby visits to Angola may expose the nonimmunised to the disease (Zagheni et al., 2008; Held et al., 2005). In addition, using the multi-step approach, high-risk areas can be identified in Hardap and Ohangwena regions. Regular surveillance of population movement may assist in controlling the risk of the disease, particularly regular border checks and targeted vaccination of children in the areas identified as high risk or along all border areas.

Furthermore, results showed that covariates like the previous counts of measles, standardised birth rates and unemployment rates were associated with increased measles risk. Such findings are similar to what was obtained elsewhere, and they are typical of contagious diseases like measles (Doungmo et al., 2014; Araújo et al., 2005; Zagheni et al., 2008; Jasem et al., 2012; Mayet et al., 2013; Ma et al., 2014), and confirm that contact is critical at sustaining transmission and that large households are at increased risk (Araújo et al., 2005; Ma et al., 2014). Existing reservoirs of the disease are a major source for maintaining the transmission to the subsequent year. Any surveillance programme should try to eliminate as much as possible any putative source of transmission. Further, there is clear evidence that poor households and neighbourhoods, as measured by unemployment rate, are the most vulnerable. Pathways of transmission are not quite clear within poor households and neighbourhoods, but they may reflect heightened contacts of the infected and the susceptible, thus fuelling multiple infections.

We also found that vaccination reduced the risk of the disease. In fact, vaccination coverage is reported to be 90% among children aged 12−23 months countrywide (MoHSS, 2013). It is therefore imperative that existing policies such as supplementary vaccination campaign every three years or in cases of measles outbreaks in Namibia should be maintained. Furthermore, Namibia may need to improve the delivery of measles vaccines by for example borrowing and improving the standard protocol of systematic reminder/recall interventions by telephone or post, which has been proven to be an effective strategy in increasing measles vaccination coverage (Filia et al., 2013). Otherwise, improper vaccination procedures and any other vaccine-related factors may cause the resurgence of measles (Jasem et al., 2012). The failure to vaccine all susceptible persons remains an obstacle to measles elimination, as studies have shown that the measles virus can still travel along the chains of transmission among vaccinated persons and infect unvaccinated people or people who have not acquired immunity by recovering from the disease (Ma et al., 2014).

Studies have further shown that care-seeking patterns are among the reasons of missing measles vaccination and the under-reporting of measles cases (Jasem et al., 2012; Filia et al., 2013). Although the existence of factors that may affect care-seeking patterns is acknowledged within the MoHSS in Namibia, there are no data documenting factors affecting care seeking. Thus, there is a need of a study on care-seeking patterns for measles which can inform better strategies of measles control.

A number of significant weaknesses of this study are acknowledged. Firstly, the use of aggregated data over the period of 2005−2014 would not allow the observation of any possible seasonality effects, which are quite common in infectious and contagious diseases like measles. Thus, the temporal effects were implicitly masked. Secondly, the accuracy of health information data constitutes a major concern to some extent. Currently, the HMIS database in the Ministry of Health and Social Services does not integrate data from all the MoHSS programmes and it does not routinely capture some critical child programme data (De Savigny et al., 2004). In addition, the accuracy of information may depend on the level of utilisation of health facilities, which in turn is influenced by the accessibility, perceived health service quality and health care seeking behaviour among many other factors (Adika et al., 2013; MoHSS, 2014a; De Savigny et al., 2004). As a result, many cases of measles are never reported to the health management system. Such under-reporting may somehow distort the geographical pattern of disease risk (Filia et al., 2013). Nevertheless, the spatial smoothing approach used in this study may have attenuated an aberrant in the measles risk spatial distribution(Doungmo et al., 2014; Lee & Sarran, 2015). Thirdly, this study has assumed that the covariates did not change in the period of ten years. Thus, the interpretation of the study findings should take into account this limitation.

In conclusion, the epidemiological implication of this study is that regional aggregated data may represent a useful data for policy and decision making at lower level, provided appropriate statistical models are developed and applied. This presents an important tool for the health sector to plan, evaluate and redesign prevention and control strategies, and make important policy decisions particularly for geographically targeted intervention in resource poor settings.

With regards to the statistical models presented here, particularly for the multi-step approach, many extensions to the fitted model are possible and they include those that can account for the temporal effects and measurement errors (Besag et al., 1991).

Chapter 4

Modelling spatio-temporal patterns of disease for spatially misaligned data: An application on measles incidence data in Namibia, 2005-2014

Background

Making inferences about measles distribution patterns at a small area (such as constituency level) is vital for more focal targeted intervention. However, in Namibia the measles data were available in aggregated format at regional level over the period 2005 to 2014. This leads to a spatial misalignment problem if the purpose is to make decisions at constituency level. Moreover, data on covariates of measles were not available each year between 2005 and 2014. Thus, assuming that covariates were constant through the study period would induce measurement errors which might have effects on the analysis results. This study presents a spatio-temporal model through a multi-step approach in order to deal with misalignment and measurement error.

Methods

For the period 2005-2014, measles data from MoHSS was analysed in two steps. First, a multi-step approach was applied to correct spatial misalignment in the data. Second, a classical measurement error model was fitted in order to account for measurement errors. The time effects were specified using a nonparametric formulation for the linear trend through first order random walk. An interaction between area and time was modelled through identically independent non-informative normal prior.

Results

The study showed that there was a high variation in measles risk distribution across constituencies and as well as over the study period (2009-2014). Furthermore, the risk of measles was found to be associated with (i) the number of people aged between 0 and 24 years, (ii) the percentage of women aged 15-49 years with an educational level more than secondary, (iii) the percentage of children aged 12-23 months who received the measles vaccine, (iv) the percentages of malnourished children under 5 years, and (vi) the measles cases for each previous year.

Conclusion

The study showed some of the determinants of measles risk and revealed areas at high risk through disease mapping. Additionally, the study showed a non-linearity temporal distribution of measles risk over the period of study. Finally, it was shown that ignoring the measurement errors may yield misleading results. It was recommended that group and geographically targeted intervention, prevention and control strategies can be tailored on the basis of these findings.

4.1 Introduction

Measles is among the most transmissible of human infections, which is caused by a virus which is a member of the genus Morbillivirus of the family of Paramyxoviridae (Bhella et al., 2007) and it is known to attack any persons, via airborne droplets, who have not had the disease or been successfully immunised (Heymann, 2015). It has an incubation period of 7 to 18 days from exposure to onset of fever (Heymann, 2015). Although the measles vaccine has been available for the past five decades, measles has remained one of the leading vaccine-preventable killer diseases among children especially in developing countries with low incomes per capita and poor health service systems (Heymann, 2015; WHO, 2014). In communities and areas where the immunisation is not widely spread, more than 90% of people are infected by the age of twenty. Because there is no antiviral treatment for the measles virus, vaccination and supportive care, such as good nutrition and adequate fluid intake, are mainly used to fight measles (WHO, 2015).

The goal of elimination of measles has been reached in countries of the Pan American Health Organisation (PAHO) through measles vaccine and careful measles surveillance. In other parts of the World Health Organisation (WHO), the complete elimination goal of measles is still to be reached with Africa and South-East Asia having set their target for 2020 (Heymann, 2015). Consequently measles cases are still reported in many countries (WHO, 2017). Various studies have shown that the distribution of measles risks vary quite often spatially due to different risk factors such as the level of immunisation, susceptible population and many other socio-economic indicators (Chiogna & Gaetan, 2004; Zhu et al., 2013). Maps resulting from spatio-temporal analysis of variations in measles incidences are often used to identify changes over time and areas of a region or a country with most disease occurrences in order to plan for a proper intervention and targeted distribution of aid to most affected areas (Zhu et al., 2013). They are indeed regarded as useful tools for geographically targeted interventions, and monitoring and evaluation of infectious diseases such as measles. However, because of confidentiality issues, spatio-temporal analyses of disease surveillance data, such as measles data, are often presented in aggregated form over time or at an area. Nevertheless, health decisions might be needed at lower administrative boundaries other than the levels where data were originally collected.

In the statistical literature, direct inferences at such levels which are made on basis of the original level of aggregation lead to a complication known as a modifiable areal unit problem (misalignment) (Finley et al., 2014). Moreover, many researchers do not account for measurement error despite the awareness of its presence and potential effects on analysis results (Buonaccorsi, 2010). Such studies assume that surrogate variables are the same as the variables of interest. Research has shown that ignoring measurement errors may, for example, lead to masking some important features of data, losing power of hypothesis testing among variables, and introducing bias in estimates (Wattanasaruch et al., 2012).

In this study, we used measles incidence data aggregated to the regional level in Namibia during 2005-2014 to fit spatio-temporal models, which would help to identify constituencies (lower level of regions) at high risk, as well as to visualise smoothed patterns of measles risk. Furthermore, the study aimed to determine factors associated with the distribution and the dynamics of measles in Namibia while accounting for measurement error that might be present in the covariates.

4.2 Methods

4.2.1 Settings

From the 2011 Namibia population and housing census (NPHC), the Namibia population stood at 2113077. Due to the presence of the arid Namib Desert, the population densities vary substantially among the regions with about more than two-thirds of the population estimated to live in the northern regions whereas less than one-tenth lives in the south.

4.2.2 Data

Data on reported measles cases over contiguous regions in Namibia are available from the Ministry of Health and Social Services (MoHSS) database for the period 2001 to 2014. Due to the improvement of the Namibia surveillance health system, the period of 2005-2014 provided consistent information for the entire country and hence only data from this period were considered in this study. The database included all suspected measles cases from which confirmed cases were extracted. A suspected case is any person with fever and maculopapular generalised rash and cough or red eyes. Whereas a confirmed case is any suspected case with laboratory confirmation or epidemiological links to confirmed cases in any outbreak (Heymann, 2015). In this study, the determination of a measles case followed the WHO standard definition, which considers a measles case as either a clinically confirmed case or an epidemiological linked case or a laboratory confirmed case (Heymann, 2015).

The following variables were considered as covariates in the model, each measured at constituency level.

(i) Number of people aged between 0 and 24 years, which represents the proxy of the size of susceptible group (Chiogna & Gaetan, 2004) for each constituency,

(ii) Employment rates,

(iii) Percentages of children aged 12-23 months who received measles vaccine (Vaccination coverage),

(iv) Educational attainment of female household population (percent of women aged 15-49 years with an educational level more than secondary),

(v) Percentages of malnourished children under 5 years,

(vi) Measles cases for each previous year were treated as the determinant factor of the subsequent year.

Table 1 gives the description of the variables used for this study. Administrative boundary maps were obtained from the Namibia Statistics Agency head office.

Table 4.1: Description of variables considered for the analysis

Variable	Description	Min,max	Source
Edu	Percentage of women aged between 15-49 years with an education more than secondary	2.7; 24.4	2013 NDHS
PrevCase	Count of measles for previous year	0; 207	MoHSS
EmployR	Employment rates	28; 92.9	2011 NPHC
LST	Number of people aged between 0 and 24 years	2691; 26605	2011 NPHC
Vacc	Percentages of children aged 12-23 months who received measles vaccine	75; 98.7	2013 NDHS
Malnou	Percentages of children under 5 years classified as malnourished according to anthropometric index of nutritional status (weight-for age: $\%below-3SD$)	0.9; 5.7	2013 NDHS

4.2.3 Statistical methods

The counts of measles cases were available at regional level and the aim of this study was to estimate the relative risk of measles at constituency level. If the data are to be analysed at regional level with the purpose of making decisions at constituency level a misalignment is introduced in the analysis. To overcome misalignment, a multi-step approach discussed in Ntirampeba, Neema, & Kazembe (2017) was used. Briefly, the multi-step approach fundamentally involves two steps. First, a total count of measles cases for constituency i is computed using the population proportional allocation of cases. Thereafter, the hierarchical smoothing techniques are used to estimate the relative risk of measles. In this study, the number of measles cases in the constituency i of region k in year j was computed as $y_{ikj} = \frac{P_{ikj}}{P_{kj}} Y_{kj}$, where Y_{kj} is the number of measles cases in the k^{th} region for year j that contains the constituency i; P_{ikj} is the total population of constituency i in the region k during year j; and P_{kj} is the total population of the region k in year j.

A Poisson hierarchical regression model was used to estimate the spatial and temporal dynamics of measles in Namibia. Thus, the following distribution for the computed measles cases was specified

$$y_{ikj} \mid m_{ikj} \sim Poisson(E_{ikj} m_{ikj}), \tag{4.1}$$

where E_{ikj} is the expected disease count for constituency i in the k^{th} region for year j, and m_{ikj} is the relative risk for contracting the disease in the constituency i of region k during the year j.

The focus in the analysis is on the form of the regression model for the log relative risk (m_{ikj}), which is specified as a function of fixed effects (i.e. covariates, where some of the covariates might not be directly observed, in the constituency i for year j), spatial random effects, temporal effects, and spatio-temporal interaction effects.

4.2.3.1 Fixed effects modelling

The fixed effects were modeled as a linear combination of covariates available in constituencies for each year. That is $X_{ikj}^T\beta$, where for fixed effect parameters β, a weakly informative Gaussian priors $\beta \sim\ N(0, \tau_\beta^{-1} I)$ with small precision τ_β on identity matrix were assumed. Alternatively, uniform vague priors may be assumed for β.

4.2.3.2 Spatial random effects modelling

The spatial trends were modelled as a sum of constituencies heterogeneities and spatial clustering effects. For the wide constituency heterogeneity (unstructured spatial random effects), ϕ_{ikj}, an independent and identically distributed prior (IID) was assumed such that $\phi_{ikj} \sim N(0, \frac{1}{\tau_\phi})$. This spatial random effect controls globally the extra-variability in the log relative risks or probability of success. Under this prior, the effect ϕ_{ikj} for each constituency is independent of all other constituencies. For the structured spatial random effects, ω_{ikj}, we assumed a Besag-York-Mollie specification (Besag & Green, 1993) such that ω_{ikj} is modelled using an intrinsic conditional autoregressive structure model (ICAR).

$$\omega_{ijk} \mid \omega_{\acute{i}jk \neq ikj} \sim N(\frac{1}{N_i} \sum_i \omega_{ikj}, \frac{1}{\tau_{\omega_i}}), \tag{4.2}$$

where τ_{ω_i} and N_i are the precision parameter and the number of neighbours of constituency i. Under this prior, the effect of ω_{ijk} for each constituency is normally distributed with mean effect equals the average of effects of neighbours of constituency i and τ_{ω_i} precision. With this model, the adjacency matrix was used to characterise the spatial relationships between constituencies. The neighbours are defined in terms of constituencies sharing at least one point (queen adjacency) and the weight is set to be one if two constituencies are neighbours, otherwise the weight equals zero (Besag & Green, 1993). The priors for the precisions of both unstructured and structured spatial random effects were assumed to be non-informative gamma distributions.

4.2.3.3 Temporal and time-space interaction effects modelling

The time effects can be modelled using time as a categorical variable through the introduction of dummy variables; using cubic splines (Dwyer-Lindgren et al., 2014); using parametric linear trends and using nonparametric formulations to relax the assumption of linear trends through random walk models (Dwyer-Lindgren et al., 2014; Blangiardo et al., 2013). In this study, we opted to specify the time effects using a nonparametric formulation for the linear trend through first order random walk and a Gaussian exchangeable prior.

$$\gamma_t \mid \gamma_{t-1} \sim N(\gamma_{t-1}, \sigma^2_{\gamma}), \quad for \quad t = 1 \tag{4.3}$$

$$\theta_t \sim N(0, \frac{1}{\tau_\theta}), \tag{4.4}$$

where γ_t and θ_t represent structured (through neighbourhood structure) and unstructured temporal effects, respectively. An interaction between area and time is modelled by expanding the temporal effects through the addition of an interaction term (δ_{it}). This interaction term explains the differences in time trend for different areas (i.e. constituencies). There exists various specifications for this term (Blangiardo, Cameletti, Baio, & Rue, 2013; Restrepo, Baker, & Clements, 2014). In this study, an identically independent non-informative normal prior was used.

$$\delta_{it} \sim N(0, \sigma^2_{\delta}) \tag{4.5}$$

Non-informative gamma prior for σ^2_{γ} and σ^2_{δ} were assumed. By combining fixed effects, spatial effects, temporal effects, and space-time interaction effects together, we obtained the regression model for the log relative risk as shown.

$$log(m_{ikj}) = log(E_{ikj}) + X^T_{ikj}\beta + \phi_{ikj} + \omega_{ikj} + \gamma_t + \theta_t + \delta_{it} \tag{4.6}$$

4.2.3.4 Measurement error models

Fundamentally, the specification of a measurement error model is based on an assumption about the distribution of the observed values given the true values or vice versa (Buonaccorsi, 2010). For the classical measurement error model, the distribution of the observed values given the true values is specified, while for the latter specification is referred to as the Berkson error model. That is, the classical measurement error model is expressed as $P(W = w \mid x)$, while the Berkson error model is given by $P(X = x \mid w)$, where X and W are the true and observed covariates, respectively. In this study, the errors in covariates were modeled using an additive non-differential classical measurement error model with respect to the response variable. In other words, the measurement error model does not depend on the value of the response variable and $w \mid x = x + u$. In this case, w are observed values of the true but unobserved covariates X (i.e. Ws are surrogate of Xs). The error term u assumed a Gaussian prior with a zero mean and a covariance matrix $C = \tau_u D(i.e.u\ N(0, C))$, where τ_u is the precision of the error term and D is a diagonal matrix of fixed scaling values (d_i) of the observational precision. By including the error model in the Eq. 4.6, the regression model for the log relative risk becomes

$$log(m_{ikj}) = log(E_{ikj}) + X_{ikj}^T\beta + W_{ikj}^T\tilde{\beta} + \phi_{ikj} + \omega_{ikj} + \gamma_t + \theta_t + \delta_{it}, \qquad (4.7)$$

where $W_{ikj}^T = \tilde{X}_{ijk} + u$ is a vector of adjusted mismeasured covariates obtained by applying a classical measurement error model on $\tilde{X}_{ijk}$ (observed mismeasured values); and $\tilde{\beta}$ is the vector of corresponding parameters. Details on measurement error models can be found elsewhere (e.g. Buonaccorsi (2010); Gustafson (2004)).

4.2.4 Analysis of measles data

A preliminary descriptive analysis of confirmed measles cases was performed to gain insight about the shifts of measles' yearly incidence (Fig. 4.1). Poisson models (Table 4.2) were built in Bayesian modelling framework using R-INLA. The first three models assumed spatial random components as the only sources of variability in the risk of measles. In these models, unstructured and structured random effects were considered.

For the unstructured random effects model (Model 1), the spatial trend includes II D random effects. Two models for the structured random effect for constituencies were considered. Model 2 assumes for each region a spatial random effect that is distributed as a function of the mean effect of regions in neighbourhood (ICAR) and Model 3 is a convolution model that assumes for each region two components of random effect, namely, specific region random effect (specific region heterogeneity) and structured random effect (random effect due to clustering). These models were extended by adding covariates and spatio-temporal component in parametric formulation fashion that assumes linearity in the global time effect and the differential trend for constituency and time (Models 4-11). To relax the assumption of linearity in constituency-time component, a non-parametric model was employed. In addition, error models were used for some variables in order to correct for mismeasuring. All models fitted in this study are summarised in Table 4.2.

The best model was selected using the deviance information criterion (DIC) given by $DIC = D + 2p$, where D is the deviance evaluated at the posterior mean and p the effective number of parameters in the model. The rule of thumb indicates that the best model is one with the smallest value of DIC. The summary statistics of the best model is presented in Table 4.3.

4.3 Results

4.3.1 Exploratory results

A total of 9923 cases were recorded for the period 2005-2014 in 13 regions in Namibia. Low records were observed for the first five years. Fig. 4.1 presents the spatio-temporal distribution of measles incidence rates in Namibia for the period 2005-2014. The regional distributions of measles incidence rates for each are shown. From Fig. 4.1, it appeared that there existed great variation in measles occurrence over the 13 regions and as well as over the study period. In four regions, namely Kavango, Khomas, Kunene, and Ohangwena, high measles incidence rates were observed through the study period. The regional and temporal variability in measles occurrence depicted by this figure has motivated the spatial-temporal analysis undertaken in this study.

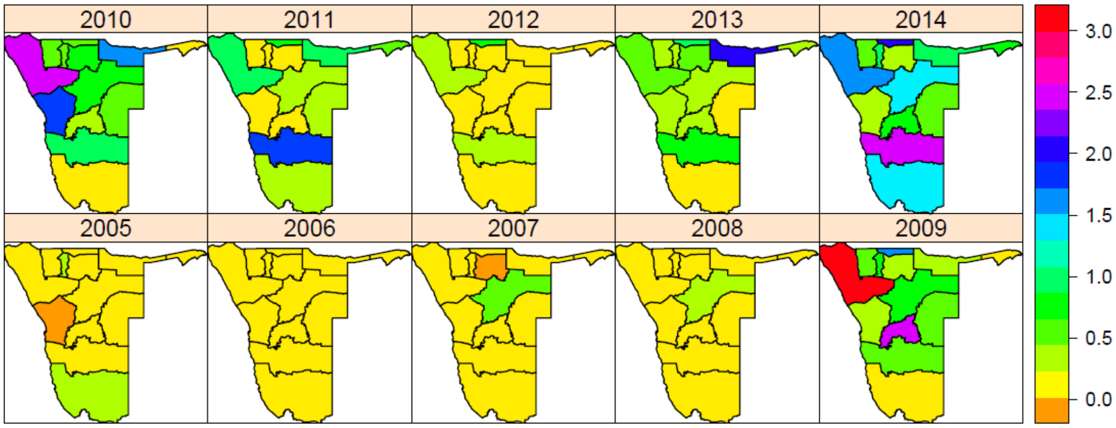

Figure 4.1: Measles incidence rates in Namibia for the period 2005-2014

4.3.2 Fixed effects

Table 4.2: Summary of models fitted to measles data for Namibia and their corresponding DICs

Poisson model	Spatial component	fixed effects	D	p	DIC
1	ICAR	-	14617.09	94.73	14806.55
2	CAR	-	14616.81	94.96	14806.73
3	IID	-	14613.48	99.71	14812.9
4	ICAR	All covariates	13169.10	152.88	13474.86
5	CAR	All covariates	13168.99	152.99	13474.97
6	IID	All covariates	13170.76	166.18	13503.12
7	ICAR	All covariates	7528.99	104.50	7737.99
8	CAR	All covariates	7528.9	104.55	7528.9
9	CAR	All covariates +time-space interaction	3767.06	619.18	5005.42
10	ICAR	All covariates +time-space interaction	3765.67	620.5	5005.77
11	CAR	All covariates +time-space interaction +measurement error	3765.54	619.21	5000.88

Based on DIC values, Model 9, which included all covariates, unstructured and structured random effects, and time-area interaction term, emerged the best fit among fitted naïve models for this data (Table 4.2). By including error models in the covariates (LST and Malnou), the Model 11 performed better than Model 9. Thus, a summary of results of this model is presented in Table 4.3. Based on the 95 % credible interval, the percent of women aged 15-49 years with an educational level more than secondary, the number of people aged between 0 and 24 years, the percentages of children aged 12-23 months who received measles vaccine, and the counts of measles for previous year, the employment rates, and the percent of children under 5 years classified as malnourished according to anthropometric index of nutritional status (weight-for age: % below $-3SD$) had significant effects on measles risks as their associated 95 % credible intervals for their fixed effects do not contain zeros.

Most of the variables are percentages except LST and LPrev variables. These two variables were transformed using natural logarithm. We performed a sensitivity

Table 4.3: Summary statistics: fixed effects (posterior mean), posterior standard deviation and posterior 95 % credible interval for Model 11

Variable	Mean	Standard deviation	95 % CI
Percentage of women aged between 15-49 with an education more than secondary	-0.0646	0.0145	-0.0935, -0.0366
Count of measles for previous year	0.1020	0.0440	0.0178, 0.1908
Employment rates	-0.0561	0.0023	-0.0616, 0.0513
Number of people aged between 0 and 24 years	0.8487	0.0727	0.7081, 0.9934
Percentages of children aged 12-23 months who received measles vaccine	-0.0379	0.0147	-0.0677,-0.0103
Percentages of children under age 5 classified as malnourished according to anthropometric index of nutritional status (weight-for age: $\%below - 3SD$)	0.0643	0.0285	0.0019, 0.1128

analysis, by fitting a model with out the variable LST, to check whether the coefficients might change significantly. Model 11 without LPrevCase variable performed equally well with the model will all parameters (DIC=5000.68). Most of the coefficients of different variables did not change significantly(except Malnou).
Negative and positive fixed effects, if exponentiated, are interpreted as decreases and increases in relative risks, respectively. For example, an increase of 1% in the percent of women aged 15-49 years with an educational level more than secondary implies a decrease of approximately 6% in the risk of measles. Also, an increase in 1 unit the log of the number of people aged between 0 and 24 years (LST) is associated with an increase of around 133.7 % in the risk of measles.

4.3.3 Spatio-temporal distribution of measles relative risks

Fig. 4.2 shows the maps of the distribution of posterior means of structured random effects, significant observed structured random effects, uncertainty around the spatial random estimates, and posterior probabilities of constituencies with specific relative risks exceeding one. For the maps of posterior means of structured random

Table 4.4: Summary statistics: fixed effects (posterior mean), posterior standard deviation and posterior 95 % credible interval for Model 11 when variable LPrevCase removed from the model

Variable	Mean	Standard deviation	95 % CI
Percentage of women aged between 15-49 year with an education more than secondary	-0.0702	0.0152	-0.0.1005, -0.0406
Employment rates	-0.0570	0.002	-0.0569, 0.0516
Number of people aged between 0 and 24 years	0.9194	0.0703	0.7812, 1.0579
Percentages of children aged 12-23 months who received measles vaccine	-0.0448	0.0146	-0.0740,-0.0166
Percentages of children under 5 years classified as malnourished according to anthropometric index of nutritional status (weight-for age: $\% below - 3SD$)	0.0516	0.0261	-0.0047,0.097

effect (a), the colours ranged from light grey to dark grey with the extreme negative random effects corresponding to extreme light grey and the extreme positive random effects corresponding to extreme dark grey. Three different numbers were used to distinguish significant observed random effects. Light grey denoted by (-1) indicated significant negative random effects, (0) indicated non-significant random effects, and dark grey denoted by (1) represented significant positive random effects (Fig. 4.2 (b)). It was observed that constituencies in Omusati, Caprivi, Omaheke, part of Kavango, and Omaruru had significant negative random effects on measles. From Fig. 4.2 (c), the spatial estimates in the northern and central parts of Namibia are associated with high uncertainty.

Fig. 4.2 (d) shows the distribution of posterior probabilities of constituencies with specific relative risks exceeding one after adjusting for covariates. In Kunene region, Opuwo, Khorixas, and Outjo constituencies had higher measles risks. There was a higher measles risk in most constituencies in Ohangwena region. In Otjozondjupa region, Okahandja and Otjiwarongo constituencies had moderate probabilities to be classified as areas at high risk of measles. For Khomas region, the constituencies in Windhoek urban had very high measles risk compared to Windhoek rural constituency.

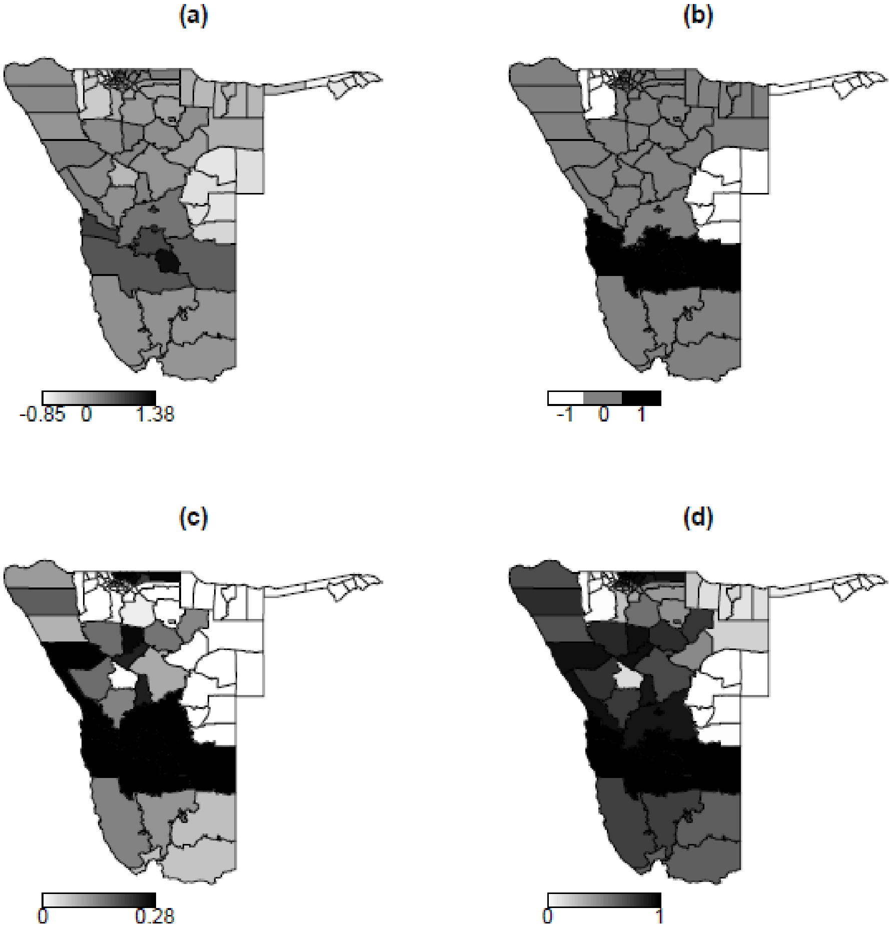

Figure 4.2: (a) map of posterior means structured random effects (Model 11) ; (b) map of significant posterior means of structured random effects (Model 11), (c) map of posterior standard error of random effects (Model 11), (d) map of posterior probabilities $p(SRR > 1 \mid y)$ (Model 11) .

All Hardap region's constituencies had high probabilities of relative risks exceeding one. The constituencies in Omusati, Omaheke, and Caprivi regions had probabilities of relative risks exceeding one close to zero.

Fig. 4.3 shows the temporal behaviour in the measles risk in Namibia between 2005 and 2014 and it concurs with temporal trend observed in measles data before smoothing techniques were applied (Fig. 4.1).

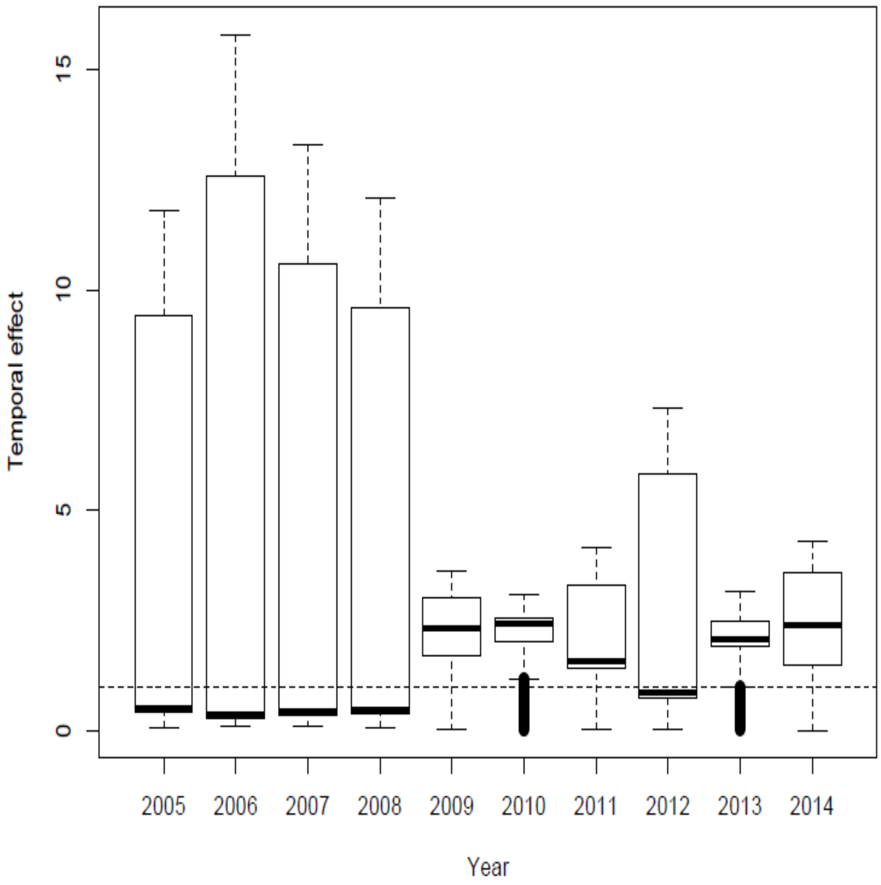

Figure 4.3: Boxplots of exponentiated posterior medians of temporal effects of measles relative risks in Namibia over the period 2005-20014.

Although there were some fluctuations in the risk of measles as the posterior means of temporal effects (when exponentiated) changed over time, it is noted that the measles risk followed an upward trend with 2009 and 2014 having remarkable peaks in measles risk (i.e. high posterior medians in temporal effect).

4.4 Discussion

The main aim in this study was to use count data that are available at regional level for the period 2005-2014 to fit an appropriate spatio-temporal model that can be used for inference at lower level (constituency level). Furthermore, the study aimed at correcting measurement errors in covariates. Thus, the study had to deal with a problem of spatial misalignment. To deal with this problem, a multi-step approach was used, which was fundamentally based on the combination of the population proportional allocation of cases for a non-uniformly distributed population and hierarchical smoothing techniques. The results of this study are consistent with previous studies that showed spatial and temporal variability in measles risk (Chiogna & Gaetan, 2004; Zhu et al., 2013; Finley et al., 2014). Like many other covariates used in this study, percentages of children under age 5 classified as malnourished and employment rates variables were only available from the 2013 NDHS. Thus, it was impossible to obtain yearly data for these covariates.

However, it would have been restrictive to assume that covariates remained constant over time. Introducing classical measurement error models in these two covariates improved the spatio-temporal ecological regression model. Model 11, which considered measurement error models, performed better than the best model (i.e. Model 9) among the naïve models (i.e. models that ignored errors in covariates). Also, results from Model 9 indicated that the percentage of children under 5 years classified as malnourished was not statistically significantly associated with the risk of measles (CI: -0.0609, 0.0772). However, when errors were accounted for in Model 11, this variable became significant (CI: 0.0019, 0.1128). This showed that indeed the dynamic of employment and nutrition had changed significantly during the period 2005-2014. In addition, this result confirmed the well-known fact that the common practice of not accounting for measurement error by the majority of researchers may yield misleading results (Buonaccorsi, 2010; Wattanasaruch et al., 2012).

This study identified the number of people aged between 0 and 24 years and the counts of measles for the previous year as significant predictors of the measles risks.

These findings are similar to the results of other studies conducted on other contagious diseases (Chiogna & Gaetan, 2004; Zhu et al., 2013; Restrepo et al., 2014; Doungmo et al., 2014) and confirmed that the proxy of social mixing and the existence of pools of the disease are critical at sustaining and continuing transmission to the subsequent years.

In this study, the percentage of women aged between 15-49 years with an educational level of more than secondary was found to be inversely associated with measles risk. This could be explained by the fact that education implies more knowledge about the risks associated with measles. In addition, education may also be considered as a proxy for social status which would imply that higher education translates to better resources and hence increased positive attitude towards health seeking (Adika et al., 2013). It is established that the attitude towards health seeking is one among other reasons for missing measles vaccination and underreporting of measles cases (Filia et al., 2013; Jasem et al., 2012).

Another finding of this study is that the percentage of children aged 12-23 months who received the measles vaccine (Vacc) are inversely associated with measles relative risks. It is therefore vital for Namibia to maintain existing policies (e.g. supplementary vaccination campaign every three years or in case of measles outbreaks) and improve the delivery of the measles vaccine by embracing strategies that are known to increase measles vaccine coverage.

The study also found the employment rate and percentage of malnourished children under 5 years to be associated with measles risks. Lower employment rates are commonly associated with poor social conditions within households, which may reflect heightened close contacts of the infected with susceptible vulnerable kids due to low nutrition. This finding concurs with results from the study by Kumar et al. (2003).

Moreover, the results showed that the constituencies in Ohangwena region were at high risk of measles. This result is consistent with previous work by Ntirampeba et al. (2017). One possible explanation is that free movement of people to and from Angola may enhance close contact of non-immunized to the disease. Indeed, Ohangwena is among the regions with a high population density and hence it has a large number of susceptible populations. These findings could be useful in designing strategies and interventions such as regular border checks and targeted vaccination in high risk or along all border areas. In addition to frequent movements of populations along the Namibian and Angolan borders, the high measles risk observed in Opuwo, Khorixas, and Outjo constituencies could be partly explained by low vaccination coverage. Although Omusati region shares borders with Angola and two regions with high risk of measles (Kunene and Ohangwena), this region is among other regions that include Caprivi, Omaheke and part of Kavango found to have a very low probability to be classified as areas at high risk. Further studies should be conducted to identify what could be the driving factors of low measles risk in these regions especially in Omusati, which seems to be an island among troubled areas. Furthermore, the study showed that Windhoek urban constituencies and all constituencies of Hardap region had very high specific relative risks of measles.

In conclusion, regional aggregated data were used to build a spatio-temporal model that is useful for constituency level inferences through a multi-step approach, while accounting for measurement errors in covariates. The study pointed out that there were significant variations in both spatial and temporal distribution of the measles occurrence in Namibia. Also, it showed factors associated with measles risks in Namibia.

On the basis of the findings of this study, we recommend the following. Firstly, the health stakeholders should increase the vaccination coverage of susceptible individuals especially in group of people aged between 0 and 24 years. Particularly, a systematic monitoring of vaccination of children aged less than five years living in poor households may help reducing the risk of measles persistence. In addition, enhancing health promotion among mothers through information, education and communication strategies should be used to improve vaccination coverage. Secondly, political leaders and stakeholders in the health sector should be able to plan and design prevention and control strategies, and make important policy decisions particularly in geographically targeted constituencies (e.g. constituencies in Kunene and Ohangwena regions). Regular surveillance of population movement may assist in controlling the risk of the disease. We particularly recommend regular border checks and targeted vaccination of children in the areas identified as high risk or along all border areas. Lastly, this study assumed that the counts of measles for the previous year (PrevCase) was not time varying. However, there might be a serial correlation between observed counts from successive years. Consequently, autoregressive models ($AR(k)$) could lead to better fits. Thus, it is recommended that future studies should expand the fitted models by handling this variable as a time varying covariate.

Chapter 5

Joint modelling spatial patterns of disease risk for data from multiple sources: An application on HIV prevalence data from antenatal sentinel and demographic and health surveys in Namibia

[2]

Background

In the disease mapping field, researchers often encounter data from multiple sources. Such data are fraught with challenges such as lack of a representative sample, which is often incomplete and most of which may have measurement errors, and may be spatially and temporally misaligned. This study presents a joint model in the effort to deal with the sampling bias and misalignment.

Methods

A joint spatial model was applied to estimate HIV prevalence using two sources: 2014 National HIV Sentinel survey among pregnant women aged 15-49 years attending antenatal care and the 2013 Namibia Demographic and Health Surveys.

[2] Published as: (Ntirampeba, D., Neema, I., & Kazembe, L. (2017)

Results

Findings revealed that health districts and constituencies in the northern part of Namibia were found to be highly associated with HIV infection. Also, the study showed that place of residence, gender, gravida, marital status, number of kids dead, wealth index, education, and condom use were significantly associated with HIV infection in Namibia.

Conclusion

This study has shown determinants of HIV infection in Namibia and has revealed areas at high risk through HIV prevalence mapping. Moreover, a joint modelling approach was used in order to deal with spatially misaligned data. Finally, it was shown that the prediction of HIV prevalence using the NDHS data source can be enhanced by jointly modelling other HIV data such as NHSS data. These findings can help Namibia to tailor national intervention strategies for specific regions and groups of population.

5.1 Introduction

Although a downwards change in the trajectory of the AIDS epidemic has been achieved worldwide (UNAIDS, 2015b), by the end of 2014, 36.9 million people were estimated to live with HIV (UNAIDS, 2015a), of which about 70 % (25.8 million) are found in sub-Sahara Africa. In 2014, it was estimated that the global total of 2 million of people were newly infected with HIV, a large portion (1.4 million) of which is said to be in sub-Sahara Africa (UNAIDS, 2015a).

Namibia is one country where the HIV prevalence is high (MoHSS, 2014b). In 2014, the number of people living with HIV among adults and children was estimated to be around 26000, of which 11000 were newly infected (MoHSS, 2014b). According to the millennium development goals (MDGs), specifically MDG6, Namibia government intended to reduce HIV prevalence among population aged 15-24 years from 8.2 % (2006) to 5 % by 2015. However, the HIV prevalence stood at 8.9 % in 2013 (NPC, 2017). Clearly, this trend points out that it would be impossible to achieve this target. From the Namibia development plan 5 (NDP 5), which is tied to the sustainable development goals (SDGs), new HIV infections per 1000 population was three for 2016/17, whereas HIV/AIDS mortality rate per 100 000 population was 134 in 2016-17 (Government, 2017). The targets for these two indicators are 1 and 90 by year 2021-2022, respectively.

The National HIV Sentinel Survey (NHSS) and Namibia Demographic and Health Survey (NDHS) are the commonly used tools to monitor the prevalence of the HIV trends in the country. Indeed, the analyses of data resulting from these surveys are vital in generating strategic information for evaluating the effectiveness of programmes and policies and enabling to improve and redesign programmes. However, each one of the two data sources has its own weaknesses that may lead to inaccurate estimations of HIV prevalence. For the former, limitations such as accessibility of ANC sites and exclusion of some categories of the population (e.g. men and non-pregnant women) are well documented (Manda et al., 2015). The latter suffers most of the times from a significant non-response drawback (Manda et al., 2015).

In the face of these limitations, a joint analysis of data from different sources has been proven to be useful (Manda et al., 2012). It avoids multiple testing on the same data, helps deal with identifiability in random effect parameters estimation, and increases precision and efficiency of parameter estimates. Further, the multivariate analysis technique can help to capture disease specific covariates and as well as to carry pairwise and cross-covariances inferences between different sources (Manda et al., 2012). Different approaches of multivariate techniques that include the multivariate normal distribution, iterative generalised least squares (IGLS) method, multivariate conditional autoregressive (MCAR) modelling, and the shared component modelling are commonly used in the mapping of multiple diseases. Although multivariate normal and IGLS methods allow modelling different sources simultaneously, these two methods underestimate the variation associated with sources (Manda et al., 2012).

In spatial disease mapping, one way to account for within and /or between areal associations is to employ the MCAR modelling approach (Manda et al., 2012). But due to high parameterisation, the computation and interpretation of parameters becomes cumbersome. Recent applications of MCAR modelling approaches include (Okango, Mwambi, Ngesa, & Achia, 2015; Gelfand & Vounatsou, 2003). Recently, shared component modelling approach pioneered by Knorr-Held & Best (2001) has been extensively used in joint analysis of multiple health outcomes (e.g. Manda, Feltbower, & Gilthorpe (2012); Manda, Masenyetse, Cai, & Meyer (2015); Knorr-Held & Best (2001); Downing, Forman, Gilthorpe, Edwards, & Manda (2008); Onicescu, Hill, Lawson, Korte, & Gillespie (2010)). This model splits the disease profile into two components, namely the disease-specific component representing spatially varying factors, and the shared component which is a proxy of unobserved spatially varying factors that are common to both diseases (Knorr-Held & Best, 2001). Bellier et al. (2013) have jointly analysed multiple data sources by including an observability parameter. Guo & Carlin (2004) have used a full Bayesian approach to link longitudinal and survival data. Other recent examples of jointly modelling multiple data sources include Bao, Raftery, & Reddy (2015), He et al. (2014), Sturrock, Pullan, Kihara, Mwandawiro, & Brooker (2013), Li, Conti, Diaz-Sanchez, Gilliland, & Thomas (2013), and Pan, Jeong, Xie, & Khodursky (2008).

Even though there is a rich literature on analyses of determinants of HIV and its geographical spread, most of the analyses used were based on univariate methods for different data sources. One notable study by Manda et al. (2015) used a shared component modelling approach to jointly analyse data from NDHS and ANC surveys. For the two sources, district level HIV prevalence rates were used and also two contextual covariates were considered as determinants of HIV. In other words, in their study, the data were first aggregated at district level and then a spatial bivariate modelling approach was applied on aggregated rates. In this situation, a misalignment in data sources was avoided. However, this has some limitations as, for instance, many covariates available from ANC or NDHS would not be used in the joint analysis. One way to include most ANC and/or NDHS covariates would be first to compute averages at district level. Alternatively, a model that allows different neighbourhood structures may be useful as it would permit to model data available at different block levels. A primary objective of this study was to develop a joint spatial model for NHSS and NDHS data, which enables the estimation at any location of the constituency or district level while dealing with misalignment in data.

5.2 Methods

5.2.1 Data

Two data sets were used in this study, namely, the 2013 Namibia Demographic and Health Survey (NDHS) data and the 2014 National HIV Sentinel Survey (NHSS) data from women aged 15-49 years attending antenatal care clinics (ANC). Table 5.1 provides a list of all variables used in this study, as identified through the literature (Manda et al., 2015; Okango et al., 2015).

5.2.1.1 NDHS data

The sampling methodology for the 2013 Namibia Demographic and Health Survey was a two stage stratified cluster survey design. In the first stage, 554 enumeration areas (EAs) were selected using probability proportional to the size of the EA, with stratification into rural and urban areas. In the second stage, 20 households were selected from each EA using equal probability systematic sampling approach. One of the key objectives of this survey was the collection of data on knowledge and prevalence of HIV/AIDS and other diseases such as diabetes, cardiovascular disease, cancer, and chronic respiratory disease (MoHSS, 2013). To achieve this objective, the survey included three questionnaires (Household questionnaire, women's questionnaire, and the men's questionnaire) that addressed questions on household characteristics and assessed women's and men's knowledge of HIV. A total of 9176 women and 3950 men formed part of the 2013 NDHS interviews. Further, the survey included HIV testing among women and men aged between 15 and 64 years selected throughout the country. Details on the survey methodologies used in collecting data can be obtained from the 2013 NDHS report (MoHSS, 2013). The variables resulting from this survey were grouped into four categories, namely, demographic, social, biological, and behavioural. The sample for the survey is thought to be a representative of the general population and also provides a vast range of population and demographic characteristics useful in the study of HIV prevalence and its related determinants.

5.2.1.2 NHSS data

Since 1992, every second year, a National HIV Sentinel survey (NHSS) has been conducted by the Ministry of Health and Social Services (MoHSS) in order to determine HIV prevalence among pregnant women aged 15-49 years attending antenatal care (ANC) clinics at public health facilities in Namibia. Since its inception, the NHSS has expanded from 8 sites to 35 district sites, supplemented by 98 satellite facilities. The main objective of the NHSS is to obtain reliable data that can be used to assess the national prevalence of HIV among pregnant women in the age group of 15-49 years; to identify socio-demographic covariates associated with high prevalence; and to fast-track the estimation of the spatial and temporal prevalence trends. Sampling techniques, sample size and data collection methods were based on the World Health Organisation (WHO) guidelines for conducting HIV surveys among pregnant women and other groups (MoHSS, 2014b). For more details, the reader can refer to the surveillance reports of the National HIV sentinel survey (MoHSS, 2014b). In this study, the 2014 NHSS, which was conducted from 10 March to 30 September 2014, was used. In total, of the 7 920 women enrolled in the 2014 NHSS, the majority of them were multi-gravida. In the data, the following variables were collected: age, gravidity, district, and HIV status. Though not many covariates are provided by the NHSS, it brings an important contribution in terms of HIV prevalence to this study as not many non-response cases are experienced in comparison to the NDHS. Table 5.1 provides a list of all variables used in this study, as identified in the literature.

5.3 Statistical models

5.3.1 Univariate modelling of data

The univariate modelling approach was achieved by fitting a separate model for each data source as follows. Let y_{ij} be a binary indicator of HIV incidence at location i (s_i) from dataset j such that y_{ij} is one if a disease incident is observed at location i for dataset j and zero elsewhere. In here, the location i could be a health district facility in a health district (for NHSS data source) or a location in a constituency (for NDHS data). Then $y_{ij} \sim Bernouilli(p_{ij})$, where p_{ij} is the probability of a recorded incident at location i from dataset j. Thus, the independent model fitted

Table 5.1: Summary of variables used in this study by source

Variable	NDHS	NHSS
1	HIV status	HIV status
2	Place of residence	Age of the respondent
3	Gender	Number of children born by a mother (Gravidity)
4	Age of the respondent	
5	Head of household	
6	Marital status	
7	Number of kids dead	
8	Education	
9	Wealth	
10	Stayed away of home	
11	Sexual activity (in last 4 months)	
12	Age at first sex	
13	condom use	
14	Had STI in last 12 months	

to dataset ($j = 1, 2$) is given by

$$logit(p_{ij}) = \beta_{0j} + \sum_{k}^{r} \beta_k X_{ijk} + f_j(g_i) + z_j(s_i), \tag{5.1}$$

with β_{0j} representing the model intercept, x_{ijk} is the k^{th} linear covariate of dataset j in a given health district facility i or constituency i, $f_j(\cdot)$ is a function of a non-linear covariate, g_i is a vector of ages, and $z_j(s_i)$ is Gaussian random field. Eq. 5.1 can be split into two separate (univariate) models as follows. At the first stage of Bayesian hierarchy,

$$logit(p_{i1}) = \beta_{01} + \sum_{k}^{r} \beta_k X_{i1k} + f_1(g_i) + z_1(s_i), \tag{5.2}$$

$$logit(p_{i2}) = \beta_{02} + \sum_{k}^{r} \beta_k X_{i2k} + f_2(g_i) + z_2(s_i), \tag{5.3}$$

For the Gaussian random field, it was assumed a multivariate Gaussian distribution $z(s) \sim N(0, \Sigma)$, where Σ is the covariance matrix. The elements of the covariance matrix Σ are specified as a function of the marginal variance of the process σ_z and the Matérn correlation function Cor_M as follows

$$\Sigma_{ij} = \sigma_z Cor_M(z(s_i), z(s_j)), \tag{5.4}$$

The Matérn correlation function is given by;

$$Cor_M(z(s_i), z(s_j)) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa \parallel s_i - s_j \parallel)^{\nu} \kappa_v(\kappa \parallel s_i - s_j \parallel), \tag{5.5}$$

where $\parallel \cdot \parallel$ denotes the Euclidean distance, $\kappa_\nu(\cdot)$ is the modified Bessel function of second order, k and ν are scale parameter and smoothness parameter respectively.

At the second stage of the Bayesian hierarchy, inverse Gamma prior distributions were assigned to k, ν, and σ_z. For fixed effect parameters β, weakly informative Gaussian priors $\beta \sim \ N(0, \tau_\beta^{-1}I)$ with small precision τ_β on identity matrix were assumed. In order to deal with non-linearity effects of continuous covariates (ages), Δg_i was assumed to follow a first order random walk process (i.e. $\Delta g_i \mid \Delta g_{i-1} \sim N(\Delta g_{i-1}, \sigma^2)$. Alternatively, a semi parametric model that uses the penalised regression spline approach may be used and details of the penalised regression approach can be found elsewhere (Okango, Mwambi, Ngesa, & Achia, 2015; Ngesa, Mwambi, & Achia, 2014).

5.3.2 Joint modelling of HIV prevalence from DHS and NHSS data sources

In the joint (bivariate) setting, the HIV prevalence from the NDHS data source and the HIV prevalence from NHSS data source were modelled jointly instead of fitting a separate model for each data source. In this study, a bivariate modelling approach was applied using the spatial shared component model that incorporated information from the NHSS source that might be common to the NDHS data source in order to improve the estimation of HIV prevalence using the NDHS source. Considering the bivariate model which pools the two datasets, let y_{ij} be a binary indicator of HIV incidence at location i from dataset $j = 1, 2$. Then $y_{ij} \sim Bernouilli(p_{ij})$, p_{ij} is the probability of recorded HIV incident pertaining to the j^{th} dataset.

The vectors relating to all observations for the two responses were concatenated in

$$\mathbf{Y} = \begin{bmatrix} y_{11} & NA \\ \vdots & \vdots \\ y_{n_1 2} & NA \\ NA & y_{21} \\ \vdots & \vdots \\ NA & y_{n_2 1} \end{bmatrix}$$

where n_ij is the number of observations for each response variable, $j = 1, 2$. Thus, the joint (bivariate) model is then given by

$$logit(p_{i1}) = \beta_{01} + \sum_{k}^{r} \beta_{k1} X_{i1k} + f_1(g_{i1}) + z_1(s_i), \tag{5.6}$$

$$logit(p_{i2}) = \beta_{02} + \sum_{k}^{r} \beta_{k2} X_{i2k} + f_2(g_{i2}) + z_2(s_i) + \gamma z_1(s_i), \tag{5.7}$$

where each response has a vector x of linear covariates with corresponding regression parameters β_{kj}; g_{ij} is the vector of ages which are assumed to follow a random walk of order 1; $z_1(s_i)$ is a Gaussian random field shared between both responses, the interaction parameter γ links the two response variables (i.e. HIV prevalence from NHSS and HIV prevalence from NDHS) and describes how much of the structure captured in $z_1(s_i)$ is also inherent in the $logit(p_{i2})$. Similar prior distributions to those specified for univariate models were assigned for parameters and hyperparameters of the joint model. A summary of models to be fitted in this study is provided in Table 5.2.

Table 5.2: Nested models to be fitted in this study

Model	GRF	Shared component	covariates
M_{U1}: Univariate model for NDHS data	$\surd$	-	-
M_{U2}: Univariate model for NHSS data	$\surd$	-	-
M_{U12}: Univariate model for NDHS data +covariates	$\surd$	-	$\surd$
M_{U22}: Univariate model for NHSS data+covariates	$\surd$	-	$\surd$
M_{J1}: Bivariate model for NDHS & NHSS data	$\surd$	$\surd$	-
M_{J2}: Bivariate model for NDHS & NHSS data+covarites	$\surd$	$\surd$	$\surd$

5.3.3 Estimation of parameters and model diagnostics

The estimation of parameters involved evaluation of the posterior distribution, which is the conditional distribution of the model parameters given the observed HIV data is obtained by taking the product of likelihood function together with the prior and hyper distributions. In this study, the posterior distribution is given by

$$p(\theta|y_{ij}) \propto \prod_{i=1}^{n} L(y_{ij}, p_{ij}) \prod_{g=1}^{2} [p(\Delta g_i|\tau_g^{-1})p(\tau_g^{-1})] \prod_{k=1}^{r} p(\beta_k)p(\tau_{\beta_k}^{-1}) \prod_{j=1}^{2} p(z_j|k_j, \nu_j, \sigma_{zk_j})p(\gamma) \tag{5.8}$$

where θ is a vector of all parameters.

A stochastic partial differential equation (SPDE) approach with R-INLA was employed to estimate posterior marginal distributions and any other posterior inferences. Convex hull meshes (Fig. 5.1) on study area were used in order to avoid the boundary effect (Krainski & Lindgren, 2013). Fig. 5.1 presents the subdivision of the domain of study into a collection of non-intersecting triangles with a condition that any two triangles meet at most a common edge or corner. The initial vertices are placed at the locations for observations and then additional vertices are added in a way that minimises the number of triangles needed to fill up the size and shape of the study domain of interest (Namibia). The polygon of triangles was extended out of the Namibian boundaries in order to avoid boundary effects. The best model was identified using the deviance information criterion (DIC) given by $D + 2p$, where D is the deviance evaluated at the posterior mean and p the effective number of parameters in the model. By the rule of thumb, the best model is one with the smallest value of DIC.

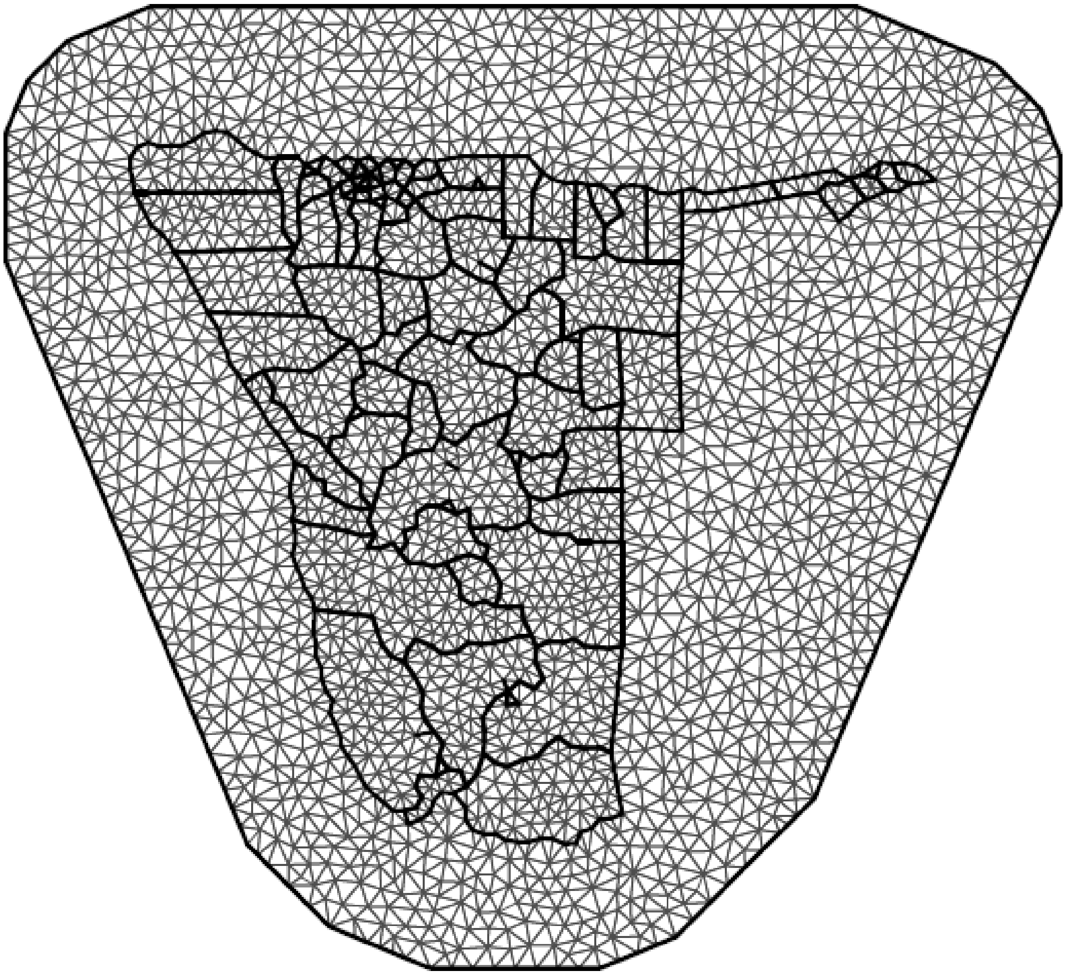

Figure 5.1: Convex hull around Namibia boundaries

5.4 Results

5.4.1 Descriptive results

Fig. 5.2(a) shows the spatial distribution of observed HIV prevalence in each constituency for women and men aged between 15 and 64 years obtained from the NDHS. This figure points out that there exist geographical (constituency level) differences of HIV prevalence in Namibia. Whereas Fig. 5.2(b) displays the geographical distribution of observed HIV prevalence among pregnant women aged 15-49 years attending antenatal care (ANC) clinics at public health facilities in Namibia (HIV

Sentinel survey data). Two colours, namely purple and blue, were used to distinguish levels of HIV prevalence. The darker the purple, the lower is the observed HIV prevalence, whereas the darker the blue, the higher is the HIV prevalence. From this figure, it can be noted that there exist spatial differences among health districts with respect to HIV prevalence. Summaries of HIV prevalence for both NDHS and NHS data sources are presented in Tables 5.3-5.6

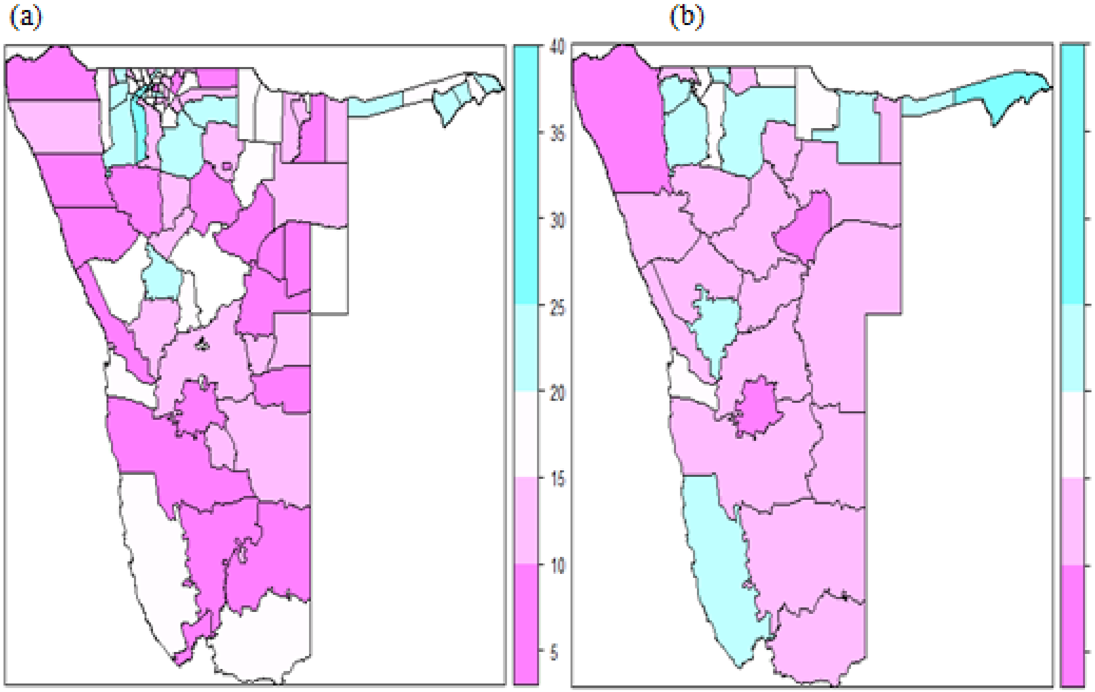

Figure 5.2: Crude HIV prevalence: (a) Constituencies HIV prevalence (2013 NDHS data); (b)Health districts prevalence (2014 NHSS data)

Table 5.3: HIV prevalence per constituency and per gender

Region	constituency	Men(%)	Women(%)	combined(%)
Caprivi	Kabe	12.9030	35.2941	24.6154
	Katima Mulilo urban	22.8571	40.2439	32.2368
	Kongola	5.2632	24.000	15.9091
	Linyati	22.222	25.000	23.8095
	Sibinda	12.000	32.000	22.000
Erongo	Arandis	6.667	7.6923	7.1429
	Daures	0	0	0
	Karibib	12.500	15.1515	13.8462
	Omaruru	17.7778	27.2727	21.7949
	Swakopmund	14.1732	14.2857	14.2276
	Walvis Bay rural	10.9091	20.833	16.5354
	Walvis Bay urban	9.7222	7.7922	8.7284
Hardap	Gibeon	4.1667	12.000	8.1633
	Mariental rural	12.5000	9.6774	10.9091
	Mariental urban	9.0909	12.9032	10.9375
	Rehoboth East urban	7.500	6.1224	6.7416
	Rehoboth rural	12.500	5.5556	8.8235
	Rehoboth West urban	0	4.5455	2.2727
Karas	Berseba	0	16.667	9.375
	Karasburg	12.5	19.048	16.22
	Keetmanshoop rural	0	6.250	3.23
	Keetmanshoop urban	12.0	14.89	13.40
	Luderitz	16.67	15.63	16.07
	Oranjemund	3.57	3.84	3.70
Kavango	Kahenge	14.29	16.44	15.65
	Kapako	13.16	8.16	10.34
	Mashare	0	8.33	3.70
	Mungu	6.25	22.5	15.28
	Mukwe	22.50	18.0	20.0
	Ndiyona	4.0	15.79	11.11
	Rundu rural east	9.375	25.49	19.277
	Rundu rural west	10.71	17.46	14.29
	Rundu urban	31.58	25	27.66

Table 5.4: HIV prevalence per constituency and per gender)(cont.)

Region	constituency	Men(%)	Women(%)	combined(%)
Khomas	Katutura central	7.27	13.41	10.95
	Katutura east	0	0	0
	Khomasdal north	4.10	3.70	3.91
	Moses Garoeb	23.30	20.86	22.325
	Samora Machel	2.84	10.78	7.64
	Soweto	0	20.75	8.94
	Tobias Hainyeko	24.02	25.0	24.43
	Windhoek east	0	0	0
	Windhoek rural	13.04	13.04	13.04
	Windhoek west	0	2.13	1.34
Kunene	Epupa	6.6670	16.67	7.69
	Kamanjab	9.52	5.88	7.89
	Khorixas	4.76	6.25	5.66
	Opuwo	10	12	11.11
	Outjo	17.86	11.43	14.29
	Sesfontein	15.38	0	7.69
Ohangwena	Eenhana	6.90	7.89	7.46
	Endola	14.29	22.50	19.67
	Engela	10.20	34.25	24.59
	Epembe	30.0	10.0	18.0
	Ohangwena	16.67	15.80	16.21
	Okongo	2.86	12.5	8.79
	Omulonga	7.14	20.0	14.13
	Omundaungilo	0	28.57	14.29
	Ondobe	2.77	24.19	16.32
	Ongenga	8.33	16.22	13.11
	Oshikango	4.08	25.97	17.46
Omaheke	Aminuis	5.26	5.56	5.41
	Epukiro	8.33	0	5.26
	Gobabis	10.52	10.0	10.26
	Kalahari	20.0	4.76	11.11
	Otjinene	5.88	4.17	4.88
	Otjimbinde	0	0	0
	Steinhausen	7.14	10.0	8.33

Table 5.5: HIV prevalence per constituency and per gender)(cont.)

Region	constituency	Men(%)	Women(%)	combined(%)
Omusati	Anamulenge	11.76	16.67	14.63
	Elim	25.0	27.78	26.47
	Etayi	14.89	15.28	15.13
	Ongongo	9.09	20.51	16.39
	Okahao	18.75	20.83	20.0
	Onesi	8.0	14.29	11.32
	Oshikuku	0	35.29	15.79
	Otamazi	20.0	29.62	26.19
	Outapi	17.39	22.67	20.66
	Ruacana	18.18	12.12	15.58
	Tsandit	9.09	30.30	20.66
Oshana	Okaku	8.7	22.86	17.24
	Okatana	7.69	36.36	20.83
	Omupunda	16.67	58.33	44.44
	Ondangwa	11.67	28.09	21.45
	Ongwendiva	8.23	10.42	9.39
	Oshakati east	18.18	22.86	20.8
	Oshakati west	16.67	18.18	17.58
	Uukwiyu	4.17	21.43	13.46
	Uuvudhiyat	11.76	-	11.76
O/shikoto	Eengondi	17.65	28.21	22.22
	Guinas	11.11	7.14	10.17
	Okankolo	5.26	22.22	10.71
	Olukonda	26.67	4.55	13.51
	Omuntele	9.09	26.47	19.64
	Omuthiyagwiipundi	19.57	27.08	23.4
	Onayena	10.52	24..32	19.64
	Oniipa	17.5	8.77	12.37
	Onyaanya	6.25	14.29	10.81
	Tsumeb	1.85	16.78	4.42
	Otjiwarongo	13.19	13.48	13.33
	Tsumukwe	12.50	12.5	12.5

Table 5.6: Observed HIV prevalence per health district/site

Heath district/site	Tested	Negative	Positive	Prevalence
Andara	255	204	51	20
Aranos	138	122	16	11.59
Eenhana	215	187	28	13.02
Engela	259	200	59	22.78
Gobabis	158	138	20	12.66
Grootfontein	222	191	31	13.96
Karasburg	214	183	31	14.49
Katima Mulilo	375	240	135	36
Keetmanshoop	163	140	23	14.11
Khorixas	180	157	23	12.78
Luderitz	278	220	58	20.86
Mariental	191	168	23	12.04
Nankudu	195	164	31	15.9
Nyangana	279	244	35	12.54
Okahandja	195	169	26	13.33
Okahao	228	181	47	20.61
Okakarara	156	142	14	8.97
Okongo	263	217	46	17.49
Omaruru	170	148	22	12.94
Onandjokwe	299	232	67	22.41
Opuwo	155	149	6	3.87
Oshakati	286	234	52	18.18
Oshikuku	306	249	57	18.63
Otjiwarongo	236	202	34	14.41
Outapi	254	225	29	11.42
Outjo	188	167	21	11.17
Rehoboth	154	140	14	9.09
Rundu	303	230	73	24.09
Swakopmund	210	188	22	10.48
Tsandi	277	221	56	20.22
Tsumeb	257	219	38	14.79
Usakos	119	93	26	21.85
Walvisbay	219	176	43	19.63
Windhoek	330	284	46	13.94
Overall	7727	6424	1303	16.86

5.4.2 Estimated HIV prevalence

Fig. 5.3(a) shows the estimated HIV prevalence within health districts using NHSS data source. From this figure, it can be deduced that in the northern part of Namibia, Katima is estimated to have the highest HIV prevalence (30 % to 35 %). Furthermore, for Andara, Rundu, Nakundu, Oshakati, Onandjokwe, Okahao, Tsandi, Outapi, Eenhana, Kongo and Engela health districts, the HIV prevalence is estimated between 15 % and 20 %. In the central west, Walvis Bay and Usakos health districts are estimated to be between 15 % and 20 % of HIV infection. In the south, the HIV prevalence is estimated be around 15 % in Luderitz. The rest of the health districts had a reduced association with HIV infection (the prevalence is estimated to be below 15 %). Fig. 5.3(b) presents the estimates of prevalence derived from NDHS data using univariate model. High HIV infection is predicted to be associated with most of the constituencies in Caprivi region (25 % to 30 %). Other constituencies with elevated HIV prevalence are found in Omusati, Oshana, Oshikoto, and Kavango regions (15 % and 20 %). Karibib, Walvis Bay rural, Walvis Bay urban, and Luderitz were estimated to have approximately between 10 % and 20 % of HIV infection. The rest of the constituencies are estimated to have HIV prevalence of around 10 %.

Fig. 5.4 provides the estimates of HIV prevalence obtained from the bivariate model that pools the two datasets together. The bivariate model reveals an under estimation of HIV prevalence when the NDHS source is used for estimation separated from the NHSS source. The univariate model estimated the prevalence to be between 0 % and 30 %, whereas the bivariate model estimated the HIV prevalence to be between 0 % and 35 %. For both data sources, the spatial distribution of HIV infection is very similar to the spatial distribution of HIV infection when univariate models were employed.

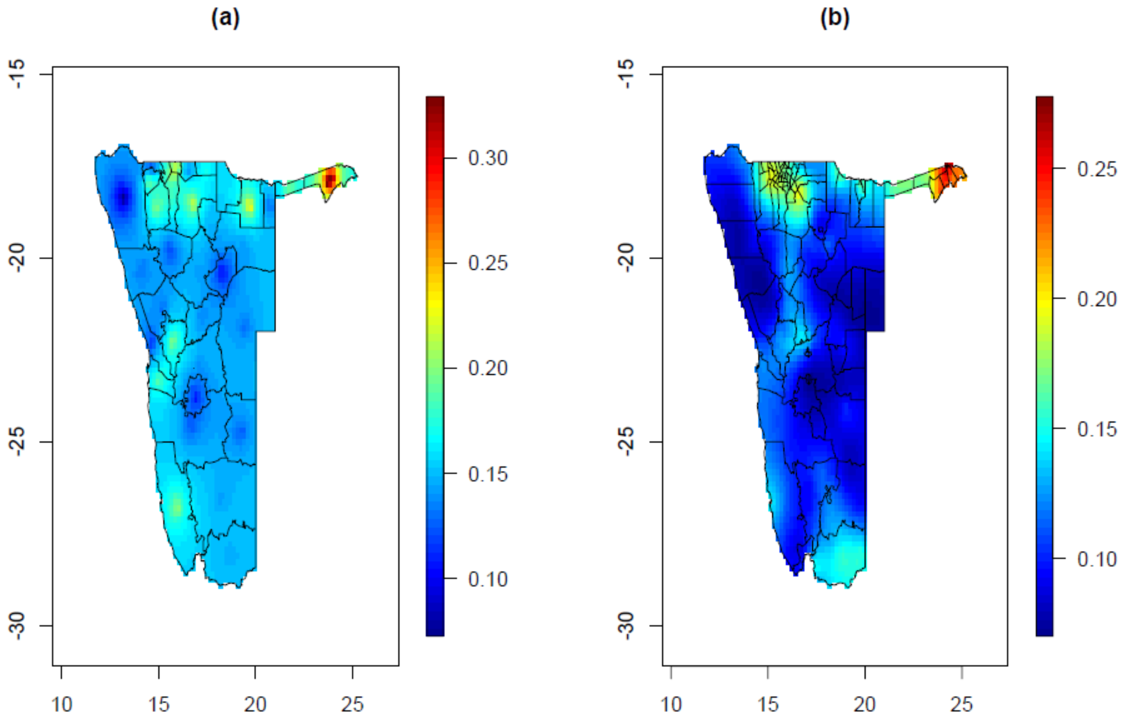

Figure 5.3: Estimated HIV prevalence using separate models: (a) HIV prevalence estimates from 2014 NHSS data; (b)HIV prevalence estimates from 2013 NDHS data

5.4.3 Linear fixed effects and nonlinear effects

From Table 5.7, it can be noticed that model M_{j2} is the best model among all models. Thus, a summary statistics of this model is presented in Tables 5.8 and 5.9 and the interpretation of the results is provided in the subsequent sections. The results of the separate (univariate) model for each data set are provided in Tables 5.11, 5.12 and 5.10, respectively.

5.4.3.1 HIV risk and its determinants: NHSS data

For the NHSS data source, two covariates namely age and gravida were available at district level (Table 5.8). The age covariate was modelled using the first order random walk in order to deal with the nonlinearity whereas the gravida covariate was assumed to have linear effects on HIV. The odds of HIV infections among pregnant women with multi-gravida (mother had given birth to two more children) was 1.88 times as likely as that of women with prima-gravida (only one child born)(OR:

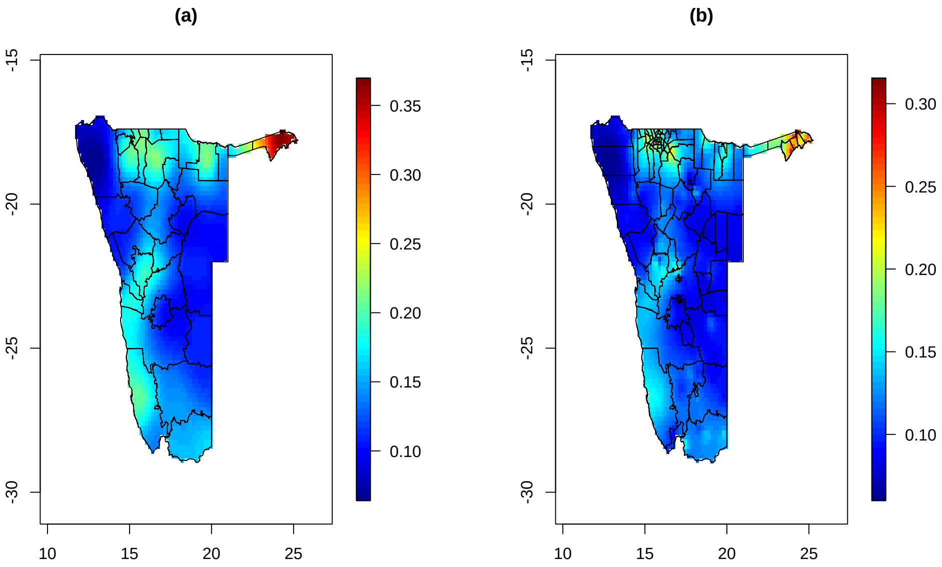

Figure 5.4: Estimated HIV prevalence using the bivariate model: (a) HIV prevalence estimates for 2014 NHSS data; (b)HIV prevalence estimates for 2013 NDHS data

Table 5.7: DIC values for fitted models

Model	DIC		
M_{U1}	7011.89		
M_{U2}: Univariate model for NHSS data	6872.59		
M_{U12}: Univariate model for NDHS data +covariates	6344.00		
M_{U22}: Univariate model for NHSS data+covariates	6388.33		
M_{J1}: Bivariate model for NDHS & NHSS data	NDHS-DIC	NHSS-DIC	Total DIC
	7003.498	6870.218	13873.72
M_{J2}: Bivariate model for NDHS & NHSS data+covarites	NDHS-DIC	NHSS-DIC	Total DIC
	5998.11	6355.98	12354.09

1.88, 95 % CI: 1.52 to 2.32). Fig. 5.5(a) shows the relationship between the age of a pregnant woman and its effects on HIV infection. This figure shows that the likelihood of HIV infection follows a nonlinear growth trajectory (black lines indicate the nonlinear trajectory whereas the dotted lines represent its 95 % credible interval). An increase in the odds of HIV infection is observed up to a certain age and then it is followed by a decline in the risk of HIV infection.

5.4.3.2 HIV risk and its determinants: NDHS data

For the NDHS data, covariates on demographic, social, sexual behaviour, and biological characteristics were available and hence used in this study. Tables 5.8 and 5.9 present the results.

Place of residence classified as rural or urban was significantly related to HIV infection among men and women. The chance of HIV infection was lower for men and women residing in rural areas compared to those residing in urban areas (OR: 1.53, 95 % CI: 1.27 to 1.84).

Gender was also found to be significantly associated with HIV infection. The likelihood of a man being infected was 0.68 times lower compared to that of a woman (95 % CI: 0.58 to 0.79).

Table 5.8: Estimated covariate effects and their 95 % credible intervals (CI)

Joint(bivariate) model		
Covariate	OR	95 % CI
β_{01}	0.12	(0.07, 0.23)
Gravida		
Prima-gravida (ref)	1.00	
Multi-gravida	1.88	(1.52, 2.32)
β_{02}	0.08	(0.04, 0.18)
Place Residence		
Rural (Ref)	1.00	
Urban	1.53	(1.27, 1.84)
Gender		
Female	1.00	
Male	0.68	(0.58, 0.79)
Head of household		
Male (Ref)	1.00	
Female	1.14	(0.97, 1.33)
Marital status		
Never in union (Ref)	1.00	
Married	0.72	(0.58, 0.89)
Living with a partner	1.41	(1.16, 1.73)
Widowed	1.46	(1.06, 2.02)
Divorced	1.07	(0.66, 1.75)
Separated	1.41	(1.04, 1.91)
Number of Kids dead		
No child died (Ref)	1.00	
One child died	1.84	(1.48, 2.29)
More than one child died	2.69	(1.84, 3.91)
Education		
No education (Ref)	1.00	
Primary	1.09	(0.87, 1.37)
Secondary	0.84	(0.66, 1.06)
Higher	0.63	(0.41, 0.96)

Table 5.9: Estimated covariate effects and their 95 % credible intervals (CI) (continued)

Joint (bivariate) model		
Covariate	OR	95% CI
Wealth index		
Poorest (Ref)	1.00	
Poorer	0.93	(0.79, 1.09)
Middle	1.10	(0.89, 1.35)
Richer	0.78	(0.61, 0.99)
Richest	0.33	(0.24, 0.46)
Stayed away from home		
Did not move away (Ref)	1	
Moved away	0.93	(0.79, 1.09)
Sexual activity		
Never had sex (Ref)	1.00	
Not active	0.98	(0.90, 1.07)
Active	1.15	(1.06, 1.26)
Age at first sex (in years)		
Never had sex (Ref)	1.00	
$\leq$ 11	1.29	(0.87, 1.91)
12 to 14	1.08	(0.67, 1.73)
15 to 17	1.47	(0.99, 2.17)
18 and above	1.26	(0.85, 1.87)
Condom used		
No(Ref)	1.00	
Yes	1.78	(1.53,2.07)
Had STI in last 12 months		
No (Ref)	1.00	
Yes	1.05	(0.96, 1.16)

The head of a household was found to be significantly linked with HIV infection. Men or women living in a household headed by a woman had higher risks of infection compared to one living in a household headed by a man (OR: 1.14, 95 % CI: 0.97 to 1.33), though not significant.

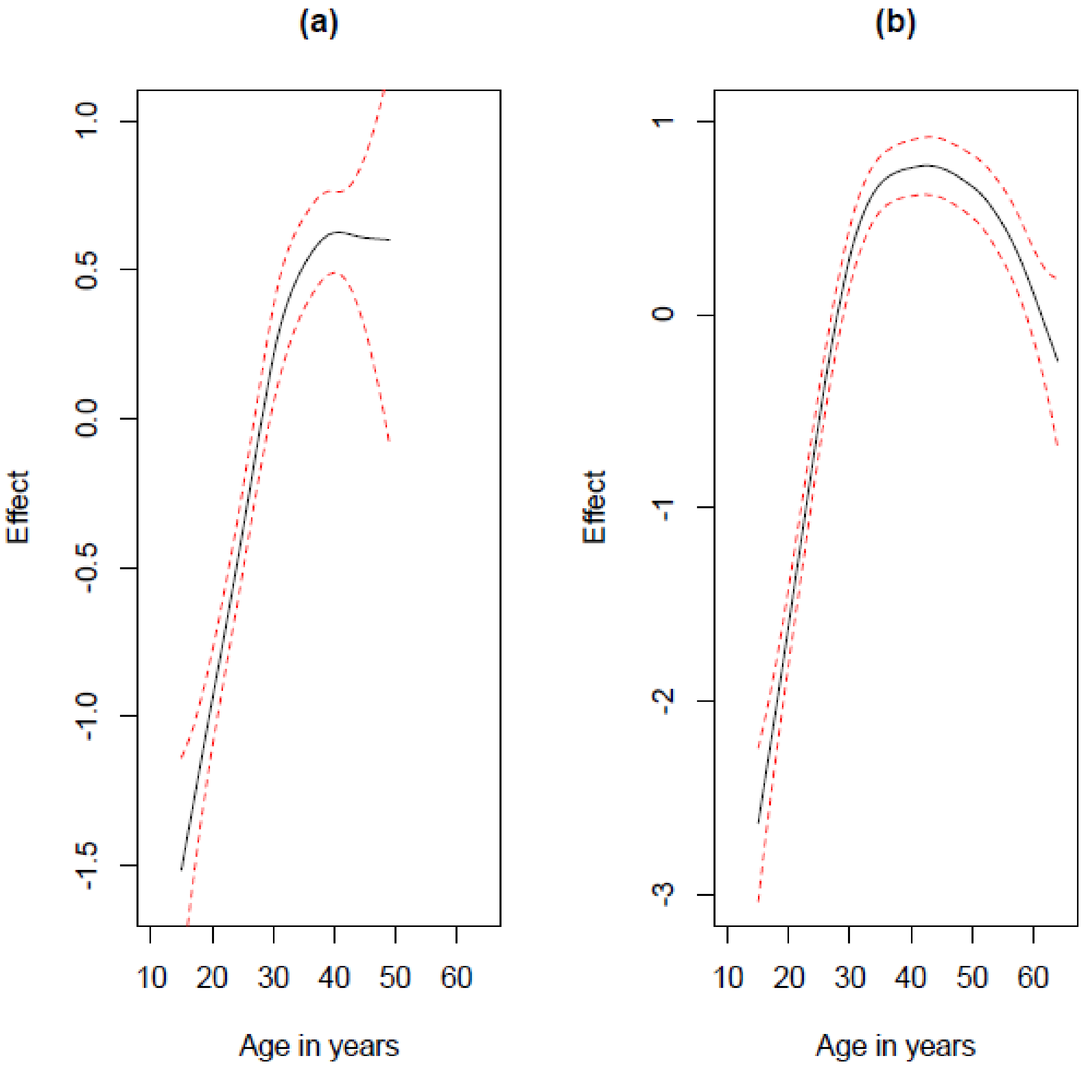

Figure 5.5: Estimated nonlinear effects of age on HIV infection and corresponding confidence intervals: (a) NHSS data; (b) NDHS data

Men and women who were married had a less risk of infection compared to those who were never in union (OR: 0.72, 95 % CI: 0.58 to 0.89). The likelihood for HIV was higher for widowers compared to men and women who were never in a union

(OR: 1.46, 95 % CI: 1.06 to 2.02). The odds of HIV infection among men and women living with partners was 1.483 times higher than that of those who were never in union (OR:1.41, 95 % CI: 1.16 to 1.73). Those who divorced had 1.07 times higher chance of infection relative to those who were never in union, though it is not significant (OR: 1.07, 95 % CI: 0.66 to 1.75). The chance of HIV infection for those who separated or non-longer lived with their partners is 1.41 times higher than that of those who were never in a union (OR:1.41, 95 % CI: 1.04 to 1.91).

The likelihood of infection with HIV for men and women who had one of their children dead is as 1.84 times higher as those whom none of their children died (OR: 1.84, 95 $\%CI$: 1.48 to 2.29). Individuals who had more than one of their children dead were 2.69 times more likely to be infected with HIV relative to those who did not have any of their children dead (OR:2.69, 95 % CI: 1.84 to 3.91).

Education was found to be negatively associated with HIV infection. The likelihood of testing positive was lower for men and women with secondary and or higher education as compared to those with no education. For instance, the odds of being infected with HIV was 0.63 times lower for men and women with higher education as compared to those with no education (OR:0.63, 95 % CI: 0.41 to 0.96).

Wealth was found to be inversely associated with HIV infection. The chance of infection with HIV was 0.78 times less for those classified as richer than that of those classified as poorest (OR:0.78, 95 % CI: 0.61 to 0.99). The men and women in the category of the richest were 0.33 times less likelihood of getting HIV as compared to those in the category of the poorest (OR: 0.33 , 95 % CI: 0.24 to 0.46).
Those in the middle class had 1.10 times odds of testing positive as compared to those in the lowest class (OR: 1.10, 95 % CI: 0.89 to 1.37), though not significant. Although not significant, individuals classified as poorer were 0.93 times less likely to test positive as compared to those classified as poorest (OR: 0.93, 95 % CI: 0.79 to 1.09).

Tables 5.8 and 5.9 also show that sexual behaviour characteristics that include current sexual activity, condom use, and age at first sex were found to be related to HIV infection. Contrary to general myth, condom use was found to be positively related to HIV infections. Individuals who ever used condoms during their last sex with most recent partners were 1.78 times at higher risk of HIV infection as compared to those who did not use condoms during their last sex with most recent partners (OR: 1.78, 95 % CI : 1.53 to 2.07).

Individuals with a history of STI in the last 12 months were 1.05 times more likely to be HIV positive relative to those who did not contract STI in the last 12 months (95 % CI: 0.96 to 1.16), though the difference is not significant.

People who had been away from their homes for more than one month in the last 12 months were found to be less likely to be HIV positive compared to those who did not go away from their homes for more than one month in last 12 months (OR: 0.93, 95 % CI: 0.79 to 1.09), although the difference was also not significant.

Fig. 5.5(b) shows that the odds of getting infected with HIV increases up to a certain age and then starts dropping at an increasing rate. This figure exhibits similar patterns to those shown in Fig. 5(a) except that the ages of respondents for Fig. 5.5(a) do not go beyond 49.

5.4.4 Spatial effects

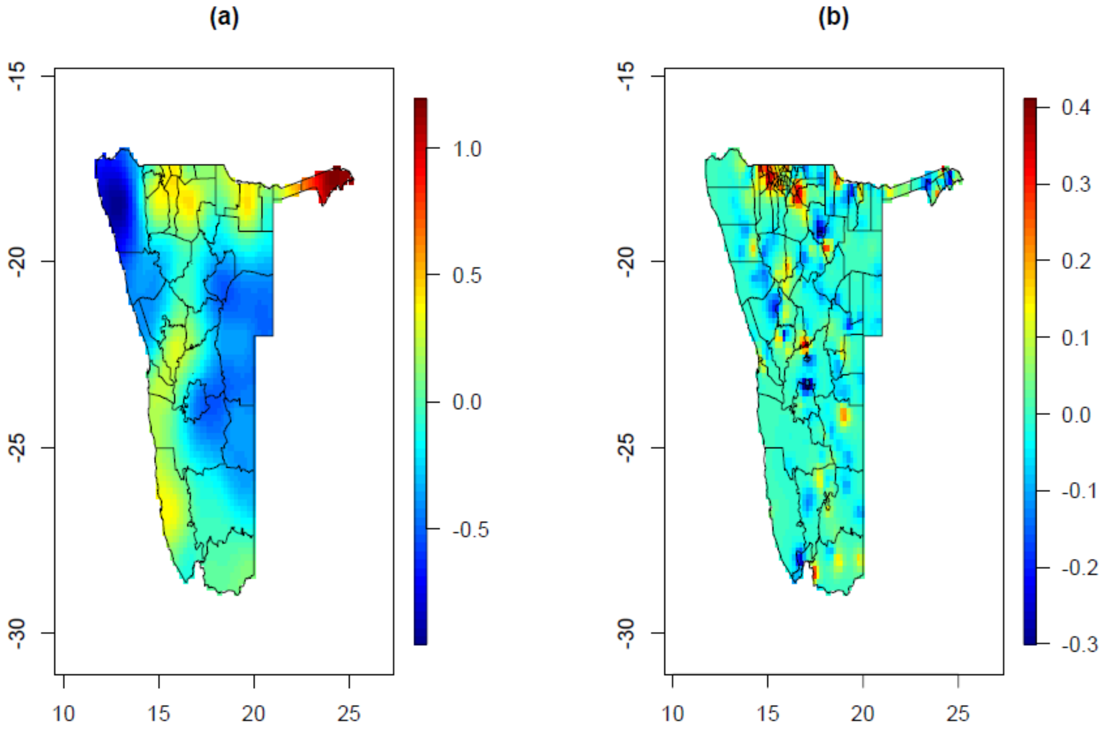

Figure 5.6: Posterior means of random effects from bivariate model: (a) Spatial random effects (NHSS data); (b) spatial random effects (NDHS) data

The maps of spatial random effects can be obtained from Fig. 5.6 and Fig. 5.7. Both figures show that health districts and constituencies in the northern part of Namibia were more likely to be associated with HIV infection (i.e. positive posterior means of spatial random effects) whereas most of the rest of the health districts and constituencies had a reduced association with HIV infection (i.e. negative posterior means spatial random effects). Fig. 5.6(a) shows the association of HIV infection with Namibia health districts. A positive posterior mean of random effects indicates a health district with a high risk of HIV infection whereas a negative posterior mean implies a relatively reduced likelihood of HIV infection. From this figure, it can be deduced that in the northern part of Namibia, Katima Mulilo, Andara, Rundu, Onandjokwe, Okahao, and Engela health districts are highly associated with HIV infection. In the central east, Walvis Bay and Usakos health districts are significantly associated with HIV infection. In the south, Luderitz is highly related to HIV infection. The rest of the health districts had a reduced association with HIV infection. Fig. 5.6(b) presents the posterior means of random effects of constituencies. The interpretation of random effects is the same as the one presented in Fig. 5.6(a). The spatial distribution of HIV infection risk in constituencies is very similar to the spatial distribution of HIV infection observed in health districts. Constituencies that are highly related with HIV infection are found in the northern part of Namibia, specifically in Caprivi, Kavango, Omusati, and Oshana, Oshikoto, and Ohangwena regions. In central east of Namibia, Walvis Bay constituency showed a moderate association with HIV infection whereas in the south Keetmanshoop urban and Oranjemund were found to be moderately related to HIV infection.

5.5 Discussion

In this study, a bivariate model controlling for spatial random effects was fitted. A full Bayesian framework through SPDE approach with INLA was implemented by jointly modelling the two data sources available at two different spatial levels. Thus, this joint model approach had to deal with data that were spatially misaligned. The bivariate model, which used a spatial shared component that acts as a surrogate of HIV risky behaviours among pregnant women in order to improve the estimation of HIV prevalence using the NDHS source, was found to be more appropriate in estimating HIV prevalence. The interaction parameter $\gamma = 2.14$ (95 % CI: 1.65 to 3.67), described how much of the structure captured in the shared component and also inherent in the NDHS HIV prevalence, was found to be significant. Hence, the joint analysis of NDHS and ANC sources has enhanced the estimation of HIV prevalence using the demographic and health survey (NDHS). This finding concurs with results from the study by Manda et al. (2015).

As everything that rises must converge (Sterman, 2000), it is argued that no quantity can grow for ever. Thus, the effect of age on HIV infection was considered to follow a growth trajectory with the two chronological patterns, namely a gradual increase from the beginning until the maximum is reached, and thereafter a gradual decrease. Consequently, it could have been inappropriate to assume that there is a linear relationship between age and the HIV infection. Therefore, in this study, the effect of age on HIV infection was modelled using first order random walk.

For these two data sources, the relationship between age and its effects on HIV infection followed an inverted U shape. This finding agrees with other studies (Okango et al., 2015; Ngesa et al., 2014).

The place of residence was found to be significantly associated with HIV infection. Individuals in urban areas had a high risk of getting infected compared those those in rural areas. This finding has been reported in many other studies (Manda et al., 2015; Okango et al., 2015; Ngesa et al., 2014; Amornkul et al., 2009). It could be used to design focused public campaigns against HIV/AIDS such as campaigns for volunteer testing and the use of condoms and antiretroviral therapy based on the place residence.

This study had shown that poverty levels were inversely associated with the likelihood of HIV infection. People in the middle class, rich class, and richest class had less risk of getting infected with HIV relative to those in the lower class. In a similar study (Chege et al., 2012), unwanted or forced sex was related to lack of resources and the ability to obtain resources.

In this study, HIV infection was found to be significantly related with the head of a household. Individuals living in a household headed by a woman were associated with higher risk of testing positive compared to the ones living in a household headed by a man. It has been shown that male-headship is a proxy of a better socio-economic status (Musenge, Vounatsou, Collinson, Tollman, & Kahn, 2013), which has been proven to be inversely related to HIV infection. This finding could be explained by the complex of inferiority of women (Mufune, Kaundjua, & Kauari, 2014) and the struggle to obtain leadership positions and power to make decisions (Chege et al., 2012).

Another finding of this study is that gender was significantly associated with HIV infection. The likelihood of women to test HIV positive is high than that of men. Some of the possible explanations for this finding are gender inequality in the sex intimacy and relationship, multiple partners perceived as prestigious for boys, and the complex of inferiority among girls in the presence of boys (Mufune et al., 2014). The gender and HIV infection relationship was confirmed in many studies (Manda, Masenyetse, Cai, & Meyer, 2015; Amornkul et al., 2009; Chege et al., 2012; Barankanira, Molinari, Niyongabo, & Laurent, 2016).

It was found that marital status impacts on HIV infection. Widowers had a high likelihood of being infected with HIV. One of the possible justifications for this finding could be that most widowers were left by partners who died of HIV. Though differences were not significant, odds of HIV infection were higher for divorced individuals and those who no-longer lived with partners compared to those who were never in a union. This result could be useful in designing strategies and interventions intended for vulnerable groups especially widowers. Some earlier works have already indicated similar results (Manda et al., 2015; Okango et al., 2015; Amornkul et al., 2009; Barankanira et al., 2016).

Another well-known finding in many studies (Okango et al., 2015; Ngesa et al., 2014; Chege et al., 2012), which was also found in this study is that education was negatively associated with HIV infection. The likelihood of testing positive was low among men and women with higher education as compared to those with no education. This could be due to the fact that most individuals with higher education are matured and aware of the danger of HIV and they are less sexually active. Though the difference is not significant, individuals with primary and secondary education were found to be at a high risk of contracting HIV as compared to those who never had any formal education. This finding could be related to limited sexual education in Namibian schools. Although life-skills programmes tailored to equip learners with sexuality knowledge are implemented in Namibian schools, it has been argued that there is no proper training provided to teachers in this matter and also that students do not take this subject seriously as it is not examinable (Mufune et al., 2014). As the Namibian government is committed to provide education to all Namibians (Government, 2002), this finding could be used by the Government to realise the need of extending free education to other phases of formal education in order to increase the number of potential individuals who will eventually achieve high education. Also, it could be used as an indicator of a need to revise the life-skills curriculum and the implementation of exams for this subject in order to encourage learners to take it seriously.

Contrary to general myths, condom use was found to be positively related to HIV infections. Individuals who hadr used a condom during the last sexual intercourse were at a higher risk of contracting HIV as compared to those who did not use a condom. This unexpected result was also reported in the work of Ngesa et al. (2014). One of the justifications provided for this finding was that men use condoms in the earlier stage of a relationship with their sexual partners and later on give up on using them.
Another possible justification to this finding could be that many of the condom users knew their HIV status (positive) and make use of condoms to protect their partners. Also, it could be that some respondents want to show-off that they are practicing safe sex, when in fact they are not. As such you find these conflicting results of high HIV prevalence and yet there is an increased uptake of condoms. So there might be issues of unsustained or inconsistent use of condoms leading to positive association with HIV prevalence.

It was also found that the number of kids already dead had a positive significant effect on HIV infection. The likelihood of getting infected with HIV for men and women who had one or more of their children dead was higher than that of those whom none of their children died. This might imply that kids could have been infected by their mothers. With respect to this outcome, the Ministry of Health and Social Services should redouble its efforts in the implementation of prevention of mother-to-child transmission of HIV/AIDS programmes until the mother to child transmission rate which was about 2 % in 2013 (MoHSS, 2014a) drops to 0 %.

With respect to sexual behaviour or biological characteristics such as sexual activity, age at first sex and STI, this study has found that these characteristics of sexual or biological behaviour are associated with HIV infection. This result could be used to identify groups with a high risk where greater efforts should be directed. In disease mapping, the identification of areas correlated with high risk proves to be useful in designing preventative and intervention strategies such as HIV testing campaigns, accessibility and use of condoms, antiretroviral treatment, and efficient budget allocation. According to the findings of this study, great efforts in terms of primary and secondary HIV interventions should be concentrated to constituencies in the northern part of Namibia.

This study made use of a shared component through the SPDE approach to analyse jointly the two sources of data and it presented two major strengths. Firstly, the joint modelling approach developed in this study allowed to combine two data sources that are available at different spatial levels in a single model. Secondly, unlike other studies that assumed a same underlying spatial process for different sources, with the bivariate model developed it is possible to specify different spatial processes (e.g. a Poisson and Bernoulli processes) through the link function. A number of significant weaknesses of this study are acknowledged. Firstly, due to confidentiality issues, the positions of HIV cases were randomly displaced in the NDHS data source. This study did not take into account the bias that might be induced by such displacements. Therefore, the interpretation of the study findings should take into account this limitation. Secondly, the missingness is quite common in NDHS and NHSS data sets. This might somehow distort the geographical distribution pattern of disease. Nevertheless, we hope that the spatial smoothing approach employed in this study might have lessened an aberrant.

5.5.1 Conclusion

This study has shown the determinants of HIV infection in Namibia and has revealed areas at high risk of HIV infection through HIV prevalence mapping. The findings from this study and the prevalence maps produced could be used by the Ministry of Health and Social Services and any other health policy makers to identify groups of people in need of HIV support and where they live in order to efficiently allocate resources that are increasingly becoming scarce. Moreover, the study used a bivariate modelling approach that helped in dealing with spatially misaligned data. Additionally, the study has shown that the prediction of HIV prevalence using the DHS data source can be enhanced by jointly modelling other HIV data.

5.6 Supplemetary results

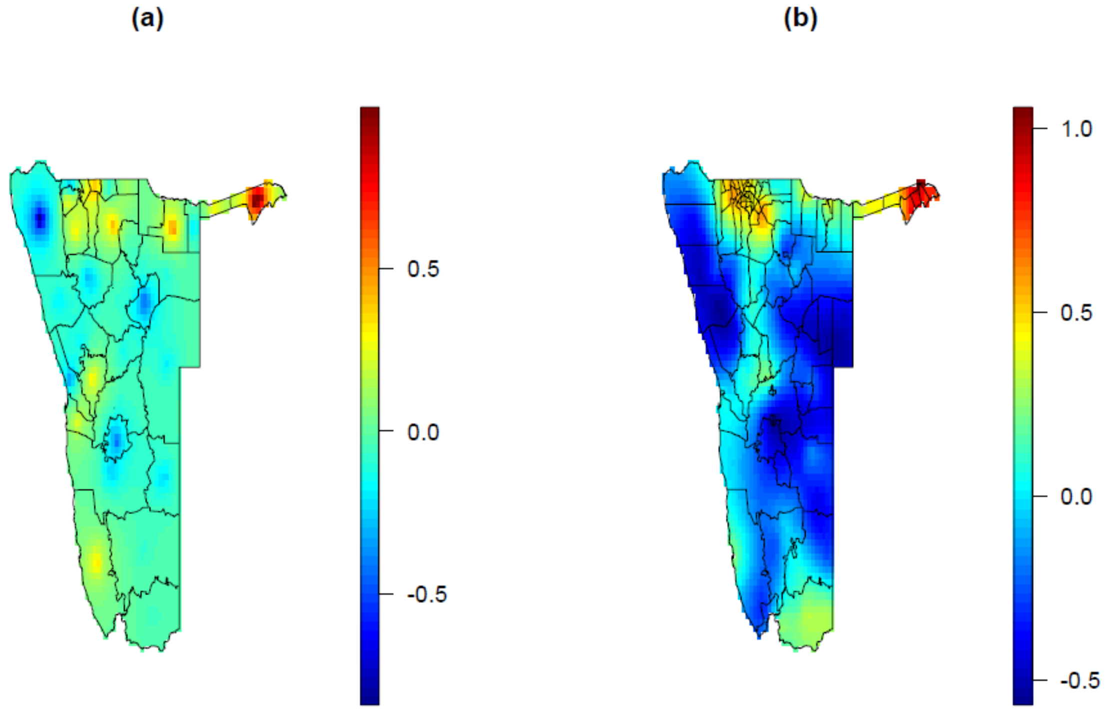

Figure 5.7: Posterior means of random effects from univariate models: (a) Spatial random effects (NHSS data); (b) spatial random effects (NDHS data)

Table 5.10: Estimated covariate effects and their 95 % credible intervals (CI): Separate model for NHSS data

Covariate	OR	95 % CI
β_{01}	0.12	(0.09,0.17)
Prima-gravida	1.00	
Multi-gravida	1.89	(1.52,2.34)

Table 5.11: Estimated covariate effects and their 95 % credible intervals (CI): Separate model for NDHS data

Covariate	OR	95 % CI
Beta	0.08	(0.03,0.21)
Place of residence		
Rural (Ref)	1.00	
Urban	1.57	(1.30, 1.89)
Head of household		
Male (Ref)	1.00	
Female	1.14	(0.97, 1.33)
Marital status		
Never in union (Ref)	1.00	
Married	0.72	(0.58, 0.89)
Living with a partner	1.43	(1.17, 1.75)
Widowed	1.49	(1.07, 2.05)
Divorced	1.09	(0.67, 1.76)
Separated	1.44	(1.06, 1.95)
Number of kids dead		
No child died (Ref)	1.00	
one child died	1.86	(1.49, 2.31)
More one than one child died	2.74	(1.88, 3.99)
Education		
No education (Ref)	1.00	
Primary	1.09	(0.87, 1.38)
Secondary	0.85	(0.67, 1.08)
Higher	0.63	(0.41, 0.96)
Wealth index		
Poorest (Ref)	1.00	
Poorer	1.1	(0.89, 1.36)
Middle	1.00	(0.80, 1.25)
Richer	0.77	(0.60, 1.00)
Richest	0.32	(0.23, 0.46)

Table 5.12: Estimated covariate effects and their 95 % credible intervals (CI): Separate model for NDHS data (Continued)

Covariate	OR	95 % CI
Stayed away from home		
Did not moved away(Re)	1.00	
Moved awayed	0.93	(0.79, 1.08)
Never had sex (Ref)	1.00	
Not active	0.98	(0.90, 1.07)
Active	1.15	(1.06, 1.26)
Age at first sex		
Never had sex(Ref)	1.00	
$\leq$ 11	1.29	(0.89, 1.96)
12 to 14	1.09	(0.68, 1.76)
15 to 17	1.49	(1.01, 2.23)
$\geq$ 18 or at first union	1.28	(0.87, 1.92)
Condom used		
No (Ref)	1.00	
Yes	1.78	(1.53, 2.07)
Had STI in last 12 months		
No (Ref)		
Yes	1.06	(0.96, 1.16)

Chapter 6

Conclusions

With the increasing availability of geographically referenced data, linking of collected data is indeed unavoidable as the exploitation of this readily available information helps in avoiding the implementation of new and expensive data collection. Proper statistical methodology is needed in order to deal with various aspects related to the analysis of geocoded data and to develop suitable statistical methods. In this dissertation, we addressed various aspects related to the analysis of spatial misaligned and mismeasured data with an inclination towards its application to measles and HIV. In this chapter, we revisit our primary objectives in order to evaluate whether they have been achieved, then we draw conclusions and make recommendations for improvements and future studies.

6.1 Review and evaluation of the objectives

- **Multi-step modelling approach in order to analysis misaligned measles data in Namibia**

With measles data, we wanted to develop a model that can be used to estimate and map the risk of measles at sub-regional level in Namibia, using data obtained at regional level. We were able to develop a multi-step modelling approach that can use data obtained at region level in order to estimate and map the risk of measles at sub-regional level in Namibia.

- **Spatio-temporal modelling while dealing with misalignment and measurement error**

The multi-step modelling approach was extended to include temporal aspect and relax the assumption of naïve analyses that assume covariates to be observed without errors.

- **Joint modelling of national HIV sentinel surveillance (NHSS) and Namibia demographic and health survey (NDHS) HIV data**

We presented a bivariate modelling approach that helped in dealing with spatially misaligned data. Additionally, we have shown that the prediction of HIV prevalence using the DHS data source can be enhanced by jointly modelling other HIV data such as NHSS data.

6.2 Lessons learnt

In this dissertation, we addressed various aspects related to the analysis of misaligned and mismeasured data and joint modelling data obtained from multiple sources with an inclination towards application to measles data and HIV prevalence from antenatal sentinel and demographic and health surveys in Namibia (i.e. NHSS and NDHS).

A mere presentation of summary statistics from statistical analyses may have little impact. A successful dissemination of results should have the purpose of informing the target audience about the findings and strategies for possible interventions. A particular strength of the disease mapping is that summary statistics can be visualized through maps which enhance an effective communication. We generated diseases risk maps for both HIV and measles, which represent important tools for the health sector to plan, evaluate and make important policy decisions particularly for geographically targeted interventions in resource poor settings.

Modelling spatial patterns of misaligned disease data

In Chapter 3, we presented a new method (i.e. multi-step method) that allows to estimate and map the risk of measles at sub-regional level in Namibia, using data obtained at regional level. This approach was applied to correct the misalignment in the data. It consisted of overlaying constituencies on regions in order to determine exactly what proportion of a given constituency is infected by measles, and applying spatial smoothing techniques to the computed measles cases to estimate disease risks. Our approach was compared with the conventionally used direct approach which is commonly used in ecology to downscale the distribution of species from coarse scale to fine scale and results showed that the multi-step approach model provided a relatively better model. Other approaches to downscale the distribution of species from coarse scale to fine scale such as hierarchical Bayesian method for interpolation, estimation and spatial smoothing (e.g. Banerjee et al. (2004); Araújo et al. (2005); Keil et al. (2013); Lee & Sarran (2015); Roli & Raggi (2015)) are commonly used. However, this class of methods relies heavily on the availability of information on a set of covariates (e.g vegetation, roads, rivers) on both grids (Roli & Raggi, 2015). In the absence of covariates information, they become impractical.

Modelling spatio-temporal patterns of disease for spatially misaligned and mismeaured data

In Chapter 4, we extended the multi-step approach by including the temporal effects and accounting for measurement errors. We introduced classical measurement error models in covariates to improve the spatio-temporal ecological regression model. Comparison of the results obtained from the naïve (i.e. modelling that ignored errors in covariates) method and those from the approach that accounts for measurement errors indicated that the latter modelling approach performed better than the former. Additionally, some covariates that were not statistically significantly associated with the risk of measles when the naïve approach was used became significant. These results led us to conclude that some other approach has to be taken in order to handle measurement errors in data.

Otherwise, analyses could produce inaccurate estimates and incorrect conclusions. The error model we presented is just one of the possible alternatives to deal with measurement errors. But, it is a useful approach to correct the measurement errors in data and improve inferences in situations where mismeasured values in covariates are encountered instead of naïve analyses.

Joint modelling of national HIV sentinel surveillance (NHSS) and Namibia demographic and health survey (NDHS) HIV data

In Chapter 5, we made use of a shared component through the SPDE approach to jointly analyse the two sources of data. Our modelling approach presented three major strengths. Firstly, the joint modelling approach developed allowed to combine two data sources that are available at different spatial levels in a single model. Secondly, unlike other studies that assumed a same underlying spatial process for different sources, with the bivariate model developed it is possible to specify different spatial processes (e.g. a Poisson and Bernoulli processes) through the link function. Thirdly, the dissertation has shown that the prediction of HIV prevalence using the DHS data source can be enhanced by jointly modelling other HIV data such as NHSS data.

6.3 Future research directions

In modelling spatial and spatio-temporal patterns of spatially misaligned measles data, old administrative boundaries, which matched existing covariates were used. However, the administrative boundaries have changed over time. Future studies could focus on using new boundaries that do not match necessarily with available covariate information as changing boundaries would introduce a misalignment between the response and covariates.

Despite clear advances in geographic information systems (GIS) and internet which make it easier to access spatial data in various forms, the ethical, policy and legal concerns still make it difficult to access the readily detailed information of individual persons. In demographic health surveys (DHS), to preserve the confidentiality of individuals, the positions of HIV cases are randomly displaced. This displacement might induce a bias in estimation and inferences. Attention to the further development of appropriate analytical methods that could facilitate to quantify and correct the impact of displacements of disease cases would add value to statistical literature as no study had looked into this issue so far. Furthermore, future work would consider extensions of the joint modelling approach presented in chapter 5. Further studies can be directed into the joint modelling of more than two data sources in order to enhance the prediction of HIV prevalence. A generalised shared component modelling within SPDE framework is a possible choice. Moreover, all models presented in chapter 3, 4, and 5 can be extended through the incorporation of a temporal misalignment component.

Appendix A

Appendix: R Codes

This appendix is subdivided into three main parts, namely, Codes of modelling spatially misaligned data :measles data, spatio-temporal modelling codes, and Joint modelling of NDHS and NHSS HIV prevalence codes. R-programmes require that all comments and other documentations are preceded by the symbol # and the commands are just statements. To avoid any confusion our codes are written within R program requirements so that these codes can be used by anyone without changing anything.

A.1 Spatial modelling data: measles data (Chapter 3)

```
#Load the package for building the map and import the shapefile
library(splines)
library(sp)
library(maptools)
library(Matrix)
library(spdep)
library(INLA)
library(foreign)
library(shapefiles)
library(BayesX)
library(lattice)
library(maps)
library(mapproj)
library(RColorBrewer)
library(latticeExtra)
library(CARBayes)
library(gridExtra)

# Load shapefile

shape = readShapePoly("G:\\Article\\Constituencies_shapefiles\\
    Old_Constituency_Boundaries_107.shp")
shape@data

# Create neighbourhood structure/adjacency matrix

neig <- poly2nb(shape)

# plot neighbourhood
plot(shape, border=gray(.5))
plot(neig, coordinates(shape), add=TRUE)

#Set up INLA call; Convert it into an inla neighbourhood object
neig.inla = nb2INLA("neig", neig) # empty variable, file is saved

#Create areas IDs used here

shape$ID<-1:nrow(shape@data)
#Read data
data<-read.csv("G:\\Article\\Measles-constituencies-covariates3.csv")

# Put the data into shape@data
old.shape = shape
shape<-data.frame(shape@data,data,shape$ID)
```

```
#Plot Moran I scatter plot, Moran I(local), and probability
 of most significant Moran I (Chapter 3, Figure 3.1)
moran.plot(data$case, listw = const_neighb,xlab="Measles cases"
, ylab="Spatially lagged measles cases")#plotting moran I
loc<-localmoran(data$case, listw = const_neighb)#Local moranIs moran I

min<-min(data$local.Ii)
max<-max(data$local.Ii)

old.shape$local.Ii=data$local.Ii
spplot(old.shape, "local.Ii",at=c(-2,0,2,4,5),main="(a)",
col.regions = rainbow(99, start=.1))

min<-min(data$p)
max<-max(data$p)

old.shape$p=data$p
spplot(old.shape, "p",at=c(0,0.05,0.5,1), main="(b)",
col.regions = rainbow(99, start=.1))
par(mfrow=c(1,2))
grid.arrange(spplot(old.shape, "local.Ii",at=c(-2,0,2,4,5),
main="(a)",col.regions = rainbow(99, start=.1))
,spplot(old.shape, "p",at=c(0,0.05,0.5,1), main="(b)",
col.regions = rainbow(99, start=.1)),ncol=2)

#Prepare the Besag model and run INLA
formula0 <- case ˜ 1+ f(shape$ID,model="besag",graph=’neig’)
besag.model <- inla(formula0,family="poisson",data=data,E=E,
    control.compute=list(dic=TRUE,cpo=TRUE))

#Preparing a model with covariates
```

```
formula1 = case~1+SAHHS+EmployR+case2004+SBrate+Vacc+
                f(shape$ID, model="besag", graph="neig")
m1 = inla(formula1, family="poisson", E=E, data=as.data.frame(shape),
          control.compute = list(config = TRUE),
          control.predictor=list(compute=TRUE))

#Calculate and map zeta (relative risk) (where csi=upsilon + nu)
m <- m1$marginals.random$‘shape$ID‘
zeta1 <- lapply(m,function(x)inla.emarginal(exp,x))
old.shape$SMR=unlist(zeta1)
spplot(old.shape, "SMR",at=c(0.4,0.8,1,2,4))

#map random effects
besag1<-m1$summary.random$‘shape$ID‘
old.shape$RE=besag1$mean
spplot(old.shape, "RE",at=c(-0.75,0,0.5,1,1.5),main="a")

#Plot Relative risk for direct method
old.shape$RR=data$RR
spplot(old.shape, "RR",at=c(0.1,1,2,3,4,11.5),main="(a)")
spplot(old.shape, "RR")

#Plot Relative risk for multi-step and direct method (chapter 3, Figure 3.2)

par(mfrow=c(1,2))
grid.arrange(spplot(old.shape, "SMR",at=c(0.4,0.8,1,2,4),main="(a)",
           col.regions = rainbow(99, start=.1))

    ,spplot(old.shape, "RR",at=c(0,1,2,6,8,11.5),main="(b)",
     col.regions = rainbow(99, start=.1)),ncol=2)

# boxplot (Figure 3.3)
boxplot(data$SresD ~ data$Method,xlab="Method",ylab="Standardized residuals")
```

```
data<-read.csv("G:\\Article\\Measles-regions-predicted_Residualsr.csv")
 #( contains response variable (Y), covariate(X), and expected cases (E)
```

A.2 Spatio-temporal modelling data: measles data (Chapter 4)

```
#0. Draw maps of regional measles incidence rates (chapter 4, Figure 4.1)

#Regions
shape = readShapePoly("G:\\Article\\Regional_boundaries\\Regional_boundaries")
shape@data
#Create areas IDs used here
shape$ID<-1:nrow(shape@data)
data<-read.csv("G:\\Article\\Measles-regions-IncidenceRate.csv")

### Put the data into shape@data
old.shape = shape
shape<-data.frame(shape@data,data,shape$ID)

old.shape$Prev1<-data$IR1
old.shape$Prev2<-data$IR2
old.shape$Prev3<-data$IR3
old.shape$Prev4<-data$IR4
old.shape$Prev5<-data$IR5
old.shape$Prev6<-data$IR6
old.shape$Prev7<-data$IR7
old.shape$Prev8<-data$IR8
old.shape$Prev9<-data$IR9
old.shape$Prev10<-data$IR10

spplot(old.shape, c("Prev1","Prev2","Prev3","Prev4","Prev5","Prev6",
"Prev7","Prev8","Prev9","Prev10")
,names.attr=c("2005","2006","2007","2008","2009","2010","2011",
"2012","2013","2014"),col.regions = rainbow(99, start=.1), layout=c(5,2))

# Model building
#1. Load  the neighbourhood structure and data
#Load  the neighbourhood structure
g="G:\\Article\\graph.dat.txt"
#Load the data
data<-read.csv("G:\\Article\\Measles-constituencies-spatio-tempo.csv")

#2. Build models without covariates
# Besag model
formula0 <- case ~ 1+ f(ID2,model="besag",graph=g)
besag.model1 <- inla(formula0,family="poisson",data=data,E=Ee,
                     control.compute=list(dic=TRUE,cpo=TRUE))
#CAR model
formula1 <- case ~ 1+ f(ID2,model="bym",graph=g)
CAR.model1 <- inla(formula1,family="poisson",data=data,E=Ee,
                   control.compute=list(dic=TRUE,cpo=TRUE))
# IID model
formula2 <- case ~ 1+ f(ID2,model="iid")
IID.model1<- inla(formula2,family="poisson",data=data,E=Ee,
                  control.compute=list(dic=TRUE,cpo=TRUE))

#3.Parametric models:  alpha + csii + (deltai + beta)*year, with covariates
formula7 <- case ~- 1+Malnou+Edu+LPrevCase+LFUnEmployR+LST+Vacc
         +f(ID2,model="iid",graph=g)+f(ID1,Year.ID1,model="rw1")
IID.parametric.model1 <- inla(formula7,family="poisson",data=data,E=Ee,
                control.compute=list(dic=TRUE,cpo=TRUE))
formula7bym <- case ~ -1+Malnou+Edu+LPrevCase+LFUnEmployR+LST+Vacc
         +f(ID2,model="bym",graph=g)+f(ID1,Year.ID1,model="rw1")
bym.parametric.model1 <- inla(formula7bym,family="poisson",data=data,E=Ee,
                control.compute=list(dic=TRUE,cpo=TRUE))
formula8 <- case ~-1+Malnou+Edu+LPrevCase+LFUnEmployR+LST+Vacc
         +f(ID2,model="besag",graph=g)+f(ID1,Year.ID1,model="rw1")
Besag.parametric.model3 <- inla(formula8,family="poisson",data=data,E=Ee,
                control.compute=list(dic=TRUE,cpo=TRUE))

#4. Build non-parametric model with no space time interaction:
 alpha + csii + gammaj + phij
#csii and are modelled through BYM
#gammaj are modelled as RW1 and rw2
#phij are modelled as exchangeable
formula9 <- case ~ 1+Edu+LPrevCase+LFUnEmployR+LST+Vacc
+f(ID2,model="bym",graph=g)+f(Year.ID1,model="rw1")+f(Year.ID2,model="iid")
BYM.nonparametric.model1 <- inla(formula9,family="poisson",data=data,E=Ee,
               control.compute=list(dic=TRUE,cpo=TRUE))

#5. Build non-parametric model with time space interaction:
alpha + csii + gammaj + phij + deltaij,
#csii are modelled through BYM
#gammaj are modelled as RW1
#phij are modelled as exchangeable
#Interaction (deltaij) is modelled as exchangeable
formula10cov <- case~-1+Malnou+Edu+LPrevCase+LFUnEmployR+LST+Vacc
    +f(ID2,model="bym",graph=g)+f(Year.ID1,model="rw1")
    +f(Year.ID2,model="iid")+f(constituency.year.ID,model="iid")

BYM.IID.nonparametric.model2cov1 <- inla(formula10cov,family="poisson",data=data,
    control.compute=list(dic=TRUE,cpo=TRUE))

#6. Create the corresponding linear combinations
lcs = inla.make.lincombs(Year.ID1= diag(10),  Year.ID2 = diag(10))

#7. Include classical error model (mec)
```

```
Malnou1=data$Malnou+runif(1070,min=0.0000001, max=0.0000002)
LFUnEmployR1=data$LFUnEmployR+runif(1070,min=0.0000001, max=0.0000002)
MEdu1=data$Edu+runif(1070,min=0.0000001, max=0.0000002)
LST1=data$LST+runif(1070,min=0.0000001, max=0.0000002)
Vacc2=data$Vacc+runif(1070,min=0.0000001, max=0.0000002)
data<-data.frame(data,Malnou1,LFUnEmployR1,MEdu1,LST1,Vacc2)
prior.beta = c(0, 0.0001)
prior.prec.u = c(1, 0.0005)
prior.prec.x = c(1, 0.0005)
prior.prec.y = c(1, 0.0005)
prec.u = 1
prec.x=1
 formula10covbE6 <- case ~+1+Edu+LPrevCase+LST+Vacc+f(ID2,model="bym",graph=g)
     +f(Year.ID1,model="rw1")+f(Year.ID2,model="iid")
     +f(constituency.year.ID,model="iid")
     +f(Malnou1, model="mec",values=Malnou1,
    hyper = list(beta = list(prior = "gaussian",param = prior.beta,
    fixed = FALSE),
prec.u = list(prior = "loggamma",
                param = prior.prec.u,
                initial6= log(prec.u),
                fixed = FALSE),
prec.x = list(prior = "loggamma",
                param = prior.prec.x,
                initial = log(prec.x),
                fixed = FALSE),
mean.x = list(prior = "gaussian",
                initial = 0,
                fixed=TRUE)))
+f(LFUnEmployR1, model="mec",values=LFUnEmployR1,hyper =
list(beta = list(prior = "gaussian",
                param = prior.beta,fixed = FALSE),
prec.u = list(prior = "loggamma",
                param = prior.prec.u,
                initial = log(prec.u),fixed = FALSE),
prec.x = list(prior = "loggamma",param = prior.prec.x,
                initial = log(prec.x),fixed = FALSE),
mean.x = list(prior = "gaussian",initial = 0,
fixed=TRUE
)
)
)
BYM.IID.nonparametric.model2covbE6 <- inla(formula10covbE6,family="poisson",
    data=data,
    list(lincomb.derived.only=TRUE), control.compute=list(dic=TRUE,cpo=TRUE))
#8. Boxplot of temporal random effects :
marginal<-lapply(BYM.IID.nonparametric. model2covbE6$marginals.lincomb.derived,
     function(X){
        marg <- inla.tmarginal(function(x) exp(x), X)
        inla.emarginal(marg)})
marginal<-unlist(marginal)
marginal<-round(marginal,4)
write.csv(marginal,file="G:\\Article\\marginal_temporal.csv") # save the marginal as
a csv file
marginal_temporalmodified<-read.csv("G:\\Article\\marginal_temporalmodified.csv")
boxplot(marginal_temporalmodified$Effect ~ marginal_temporalmodified$Year,
    xlab="Year", ylab="Temporal effect")
    abline(h=1,lty=2)

#9. Plot  maps of spatial random effects (Figure 4.4)
besag2<-BYM.IID.nonparametric. model2covbE6$summary.random$ID2
namibia<-read.bnd("C:\\Users\\Dismas\\Desktop\\Article\\namSW1.csv")
namibia<-read.bnd("G:\\Article\\namSW1.csv")
#Grey scheme (two maps in one frame)
par(mfrow=c(1,2), mar=c(3,3,.5,.5), mgp=c(1.5,.5,0), las=1)
drawmap(data=besag2,map=namibia,regionvar="ID",
plotvar="mean",swapcolors=T,cols="grey",limits=c(-0.85,1.38),density=36,
    drawnames=F,cex.legend=1,cex.names=1, main="(a)")
drawmap(data=besag2,map=namibia,regionvar="ID",
plotvar="mean",swapcolors=T,cols="grey",pcat=T,density=36,
    drawnames=F,cex.legend=1,cex.names=1, main="(b)")

#10.  Plot map of probability (Figure 4.3)
# Calculating zeta=exp(csi) where csi=upsilon + nu
m1 <- BYM.IID.nonparametric. model2covbE6$marginals.random$ID[1:107]
zeta1 <- lapply(m1,function(x)inla.emarginal(exp,x))

#Calculating the probability that the spatial effects zeta are above 1,
#identifying areas with excess risk of measles. This is equivalent to
#calculate the probability that csi is above 0,
a=0
inlaprob1<-lapply(BYM.IID.nonparametric. model2covbE6$marginals.random$ID[1:107],
                    function(X){  1-inla.pmarginal(a, X)
                    })
Spatial.results1<- data.frame(ID=seq(1,107),SMR=unlist(zeta1),pp=unlist(inlaprob1))
drawmap(data=Spatial.results1,map=namibia,regionvar="ID",
plotvar="pp",swapcolors=T,cols="grey",limits=c(0.000,1),density=36,
drawnames=F,cex.legend=1,cex.names=1)
```

A.3 Joint modelling of NDHS and NHSS HIV prevalence codes (Chapter 5)

```
# 1. Draw maps of HIV raw prevalence at district and constituency level (Figure 5.2)

###1. Load district shapefile& data
```

```
shape1 = readShapePoly("G:\\districtric_boundaries\\BDR_health districts")

### Create neighbourhood/adjacency matrix
# neig: neighbourhood structure
neig1 <- poly2nb(shape1)

# plot neighbourood
plot(shape1, border=gray(.5))
plot(neig1, coordinates(shape1), add=TRUE)

### Set up INLA call
# Convert it into an inla neighbourhood object
neig.inla = nb2INLA("neig1", neig1) # empty variable, file is saved

#Create areas IDs used here
shape1$ID<-1:nrow(shape1@data)$

data1<-read.csv("G:\\DHS2013 datasets\\District_HIV_ prevalence.csv")


### Put the data into shape@data
old.shape1 = shape1
shape1<-data.frame(shape1@data,data1,shape1$ID)


# Get mean estimates
old.shape1$prevalence = data1$Prevalence

#-----------------------------------------#

###2.load shapefile $data at constituency level##

shape2 = readShapePoly("G:\\Constituencies_shapefiles
\\Old_Constituency_Boundaries_107.shp")

### Create neighbourhood/adjacency matrix
# neig: neighbourhood structure
neig2 <- poly2nb(shape2)

# plot neighbourood
plot(shape2, border=gray(.5))
plot(neig2, coordinates(shape2), add=TRUE)

### Set up INLA call
# Convert it into an inla neighbourhood object
neig.inla = nb2INLA("neig2", neig2) # empty variable, file is saved

#Create areas IDs used here
shape2$ID<-1:nrow(shape2@data)

data2<-read.csv("G:\\DHS2013 datasets\\Const_HIV_Prevalence.csv")

### Put the data into shape@data
old.shape2 = shape2
shape2<-data.frame(shape2@data,data2,shape2$ID)

# Get mean estimates
old.shape2$CombinedPrevalence = data2$CombinedPrevalence

#----------------------------------------------------#

#3. Plot maps of raw HIV prevalences for districts
$ constitituencies in one frame
#dev.off()

par(mfrow=c(1,2), mar=c(3,3,.5,.5), mgp=c(1.5,.5,0), las=1)
```

```
grid.arrange(spplot(old.shape2, "CombinedPrevalence",
at=c(0,3,10,15,20,25,40),main="(a)") ,spplot(old.shape1, "prevalence",at=c(0,3,10,15,
,ncol=2)

# 2. Fit joint models
The execution of joint modelling is achieved in two main steps,
 namely the main functions (Functions) and codes (Codes) to execute main functions.
```

A.3.1 Functions

```
# wrapper function to create a mesh object as used in an INLA model fit
make.mesh<-function(locs = NULL, mesh.pars = NULL, spatial.polygon = NULL,
        sphere = FALSE, plot = FALSE){
        if(is.null(mesh.pars)){
        if(is.null(spatial.polygon)){
            w <- ripras(locs)
            mesh.pars <- c(max = 0.15*sqrt(area(w)) ,
                           min = 0.1*sqrt(area(w)),
                           cutoff = 0.15*sqrt(area(w)))
                }else{
                    w <- spatial.polygon
                    mesh.pars <- c(max = 0.15*sqrt(area(w)),
                                   min = 0.1*sqrt(area(w)),
                                   cutoff = 0.15*sqrt(area(w)))
                    }
        }
    # getting mesh parameters
    max.edge.min <- mesh.pars["min"]
    max.edge.max <- mesh.pars["max"]
    if(is.na(max.edge.max)){max.edge <- max.edge.min}
       else{max.edge <- c(max.edge.min,max.edge.max)}
```

```
        cutoff <- mesh.pars["cutoff"]
        mesh.pars<-c(max.edge,cutoff)
    if(!sphere){
        if(!is.null(spatial.polygon)){
            # creates triangulation based on a spatial polygon of the domain
            boundary <- inla.sp2segment(spatial.polygon)
            mesh <- inla.mesh.2d(boundary = boundary, max.edge = max.edge,
             cutoff = cutoff)
        } else {
            loc <- locs
# creates triangulation based on the locations of the point pattern
            mesh <- inla.mesh.2d(loc = locs, max.edge = max.edge, cutoff = cutoff)
        }}
    if(sphere){
        if(!is.null(spatial.polygon)){
            # creates triangulation based on a spatial polygon of the domain
             projected onto a sphere
            boundary <- inla.sp2segment(spatial.polygon)
            boundary$loc <- inla.mesh.map(boundary$loc, projection="longlat",
                                             inverse=TRUE)
            mesh <- inla.mesh.2d(boundary = boundary, max.edge = max.edge,
                                 cutoff = cutoff)
        } else {
 # creates triangulation based on the locations of the point pattern
 # projected onto a sphere
            locs <- inla.mesh.map(locs, projection="longlat", inverse=TRUE)
            mesh <- inla.mesh.2d(loc = locs, max.edge = max.edge, cutoff = cutoff)
        }}
    if(plot) plot.mesh(mesh)
    mesh
}
```

```
# function to fit a joint spatial model to geo-statistical data where one spatial
component is shared between the responses
joint.fit <- function(mesh = NULL, locs.1 = NULL, locs.2 = NULL,
response.1 = NULL, response.2 = NULL,  family = c("gaussian","gaussian"),
verbose = FALSE,
             hyper = list(theta=list(prior=’normal’, param=c(0,10))),
             control.inla=list(strategy=’gaussian’,int.strategy = ’eb’)){
    spde <-inla.spde2.matern(mesh = mesh, alpha = 2)
    # number of mesh nodes
    nv <- mesh$n
    ## create projection matrix for loacations
     Ast1 <- inla.spde.make.A(mesh = mesh, loc = locs.1)
     Ast2 <- inla.spde.make.A(mesh = mesh, loc = locs.2)
     field.1 <- field.2 <-  copy.field <-1:nv
     stk.pp <- inla.stack(tag="obs1",data=list(y=cbind(response.1,NA)),
                          A=list( Ast1,1),
                          effects=list(field.1 = field.1, beta0 =
                          rep(1,nrow(locs.1))))
    formula <- y ˜ 0 + beta0 + alpha0 + f(field.1, model=spde) +
        f(field.2, model=spde) +
        f(copy.field, copy = "field.1", fixed=FALSE, hyper = hyper )
    stk.mark <- inla.stack(tag="obs2",data=list(y=cbind(NA,response.2)),
                           A=list(Ast2, Ast2,1),
                           effects=list(field.2 = field.2, copy.field = copy.field,
                           alpha0 = rep(1,nrow(locs.2))))
    ## combine data stacks
    stack <- inla.stack(stk.pp,stk.mark)
    ##call to inla
    result <- inla(as.formula(formula), family = family,
            data=inla.stack.data(stack),
            control.predictor=list(A=inla.stack.A(stack)),
            control.inla = control.inla,
            verbose = verbose,control.compute=list(dic=TRUE))
    result
}

## function that extracts the random fields of the fitted model
find.fields <- function(x = NULL, mesh = NULL, n.t = NULL, sd = FALSE,
            plot = FALSE, spatial.polygon = NULL,...){
        if(is.null(attributes(x)$mesh) & is.null(mesh)){
        stop("no mesh has been supplied")}
    if(!is.null(attributes(x)$mesh)){mesh <- attributes(x)$mesh}else{mesh <- mesh}
    proj <- inla.mesh.projector(mesh)
    if(!is.null(spatial.polygon)) inside <- inwin(proj,as.owin(spatial.polygon))
    spde <-inla.spde2.matern(mesh = mesh, alpha = 2)
    fields <- names(x$summary.random)
    n <- length(fields)
    if(!is.null(n.t)){
        t <- n.t
        means <- list()
        for (i in 1:n){
            means [[i]] <- lapply(1:t, function(j) { r <- inla.mesh.project(proj,
            field = x$summary.random[[i]]$mean[1:spde$n.spde + (j-1)*spde$n.spde]);
            if(!is.null(spatial.polygon)) r[!inside] <- NA; return(r)})
        }
        sds <- list()
        for (i in 1:n){
            sds [[i]] <- lapply(1:t, function(j) {r <- inla.mesh.project(proj,
            field = x$summary.random[[i]]$sd[1:spde$n.spde + (j-1)*spde$n.spde]);
            if(!is.null(spatial.polygon)) r[!inside] <- NA;  return(r)})
        }
        if(!is.null(spatial.polygon)) for(i in 1:n){sds[[i]][!inside] <- NA}
        if(plot){plot.fields( x = x, mesh = mesh, n.t = n.t, sd = sd,
        spatial.polygon = spatial.polygon,...)}
    }else{
        means <- list()
    for (i in 1:n){
      means[[i]] <- inla.mesh.project(proj,x$summary.random[[i]]$mean)
      if(!is.null(spatial.polygon)) means[[i]][!inside] <- NA;
      }
    sds <- list()
    for (i in 1:n){
      sds[[i]] <- inla.mesh.project(proj,x$summary.random[[i]]$sd)
      if(!is.null(spatial.polygon)) sds[[i]][!inside] <- NA;
      }
    if(plot){plot.fields( x = x, mesh = mesh, n.t = n.t, sd = sd,
    spatial.polygon = spatial.polygon,...)}
  }
  names(means) <- names(sds) <- fields
  ifelse(sd,return(sds),return(means))
}

## find which parts of the random field are inside the supplied spatial polygon
inwin<-function(proj, window){
  e<-expand.grid(proj$x,proj$y)
  o<-inside.owin(e[,1],e[,2],window)
  o<-matrix(o,nrow=length(proj$x))
}

##function for plotting random fields called by function find.fields
plot.fields <- function(x = NULL, mesh = NULL, n.t = NULL, sd = FALSE,
   spatial.polygon = NULL,col = grey.colors(100,0.05,0.95),...){
    proj <- inla.mesh.projector(mesh)
    fields <- names(x$summary.random)
    n <- length(fields)
  par(...)
  if(!is.null(n.t)){
    rfs <- find.fields(x = x, mesh = mesh, n.t = n.t, sd = sd,
    spatial.polygon = spatial.polygon)
```

```
        t <- n.t

        for(i in 1:n){
            for(j in 1:t){ image.plot(proj$x,proj$y,rfs[[i]][[j]],
            axes=FALSE,xlab="",ylab="", main = paste(fields[i],
             "time", j,  sep = " "), col = col)
            contour(proj$x,proj$y,rfs[[i]][[j]],add=TRUE)
            }
        }
    }else{
        rfs <- find.fields(x = x, mesh = mesh, sd = sd,
        spatial.polyon = spatial.polygon)
        for(i in 1:n){
           image.plot(proj$x,proj$y,rfs[[i]],
           axes=FALSE,xlab="",ylab="", main = fields[i], col = col)
           contour(proj$x,proj$y, rfs[[i]],add=TRUE)
        }
    }
}

## function for  plotting mesh
plot.mesh <- function(x,...){
    plot(x,main="",asp=1,draw.segment = FALSE,...)
    if (!is.null(x$segm$bnd))
                lines(x$segm$bnd, x$loc, lwd = 2,col = 1)
    if (!is.null(x$segm$int))
                lines(x$segm$int, x$loc, lwd = 2,col = 1)
}
## plot for individually fitting models to geo-statistical data
geo.fit <- function(mesh = NULL,  locs = NULL, response = NULL,
    family = "gaussian",verbose = FALSE){
    # spde model for the spatial random field
    spde <- inla.spde2.matern(mesh, alpha = 2)
    # index
    index <- inla.spde.make.A(mesh = mesh, loc = locs)

    #create data stack
    stack <- inla.stack(data=list(y=response),
        A=list( index,1),
        effects=list(field = 1:mesh$n, beta0 = rep(1,nrow(locs))))
    #nl <- paste("\"",nl.model,"\"",sep="")
    formula = y~  0 + beta0 + f(field, model=spde)
    result <- inla(as.formula(formula), family = family ,
            data = inla.stack.data(stack),
            control.predictor=list(compute=TRUE, A=inla.stack.A(stack)),
            verbose = verbose,control.compute=list(dic=TRUE))
    result
}
}
```

A.3.2 Codes

```
#1. load data (dhs and hss and spatial.polygon)
library(sp)
library(maptools)
library(spatstat)
library(rgdal)
#load("HIVdata.RData")
#source model fitting function etc.
source("functions.r")
#libraries
libs <-c("INLA","fields","maptools","spatstat")
lapply(libs, require, character.only = TRUE)
## read in data for the two data sets
hss.data<-read.csv("C:\\Users\\dntirampeba\\Desktop\\DHS2013 datasets
\\HSS_2014.csv")
```

```
dis1<-hss.data
hss<-data.frame(dis1)
dhs<-read.csv("C:\\Users\\dntirampeba\\Desktop\\DHS2013 datasets\\
    MEN_WOMEN_HIV_SPDE1.csv")
dis2<-dhs
dhs<-data.frame(dis2)

## finds the name of the layer
 spatial.polygon<-readOGR("Constituencies_shapefiles\\
 Old_Constituency_Boundaries_107.shp",
 layer = "Old_Constituency_Boundaries_107")
 ogrListLayers("G:\\Constituencies_shapefiles
 \\Old_constituency_Boundaries_107.shp")

## read in the shapefile
spatial.polygonc<-readOGR("G:\\Constituencies_shapefiles
\\Old_constituency_Boundaries_107.shp",
    layer="Old_constituency_Boundaries_107")
spatial.polygond<-readOGR("G:\\districtric_boundaries\\
BDR_health districts.shp",
    layer="BDR_health districts")
#save data and spatial polygons together
save(hss,dhs,spatial.polygond,spatial.polygonc, file="DhsHss.RData")
load("DhsHss.RData")
names(dhs)

#2. Define locations: locs
loc.h<- as.matrix(dis1[,14:13])
loc.d<- as.matrix(dis2[,38:37])
##locations and response variables (note here I have whatever is suffix 2 as a copy
## of suffix1
locs.1<- as.matrix(hss[,14:13])
locs.2<- as.matrix(dhs[,38:37])
```

```
response.1 <- hss[,2]
response.2 <- dhs[,39]
##3. Choose (fixed effect) covariates for each response
fixed.covariates.1 <- data.frame(gravida = Hss[,3])
fixed.covariates.2 <- data.frame(edu1 = Dhs[,2],edu2 =Dhs[,3],edu3 =Dhs[,4],
    sexhhh=Dhs[,5],awaymonth=Dhs[,6],WealthIndexP=Dhs[,7],WealthIndexM=Dhs[,8],
    WealthIndexR=Dhs[,9],WealthIndexRt=Dhs[,10],kidsdead1=Dhs[,11],
    kidsdead2=Dhs[,12],mstatus1=Dhs[,13],mstatus2=Dhs[,14],
    mstatus3=Dhs[,15],mstatus4=Dhs[,16],mstatus5=Dhs[,17],
    agesex1=Dhs[,18],agesex2=Dhs[,19],agesex3=Dhs[,20],
    agesex4=Dhs[,21],sexualactivity1=Dhs[,22],sexualactivity2=Dhs[,23],
    condomUse=Dhs[,24],STI=Dhs[,25],placeres=Dhs[,27],
    gender = Dhs[,31])

## 4.Define nonlinear effect : c("rw1","rw1")
nl.cov.1 <- Hss[,1]
nl.cov.2 <- Dhs[,1]

##5. Visualise data
par(mfrow=c(1,2))
plot(spatial.polygond)
points(locs.1, pch = 18, col = (response.1))
plot(spatial.polygonc)
points(locs.2, pch = 13, col = (response.1+3))
legend("bottomright", legend=c("dhs HIV 0","dhs HIV 1"),pch=c(13,13),
col=c(3,4),bty="n")
## 5. Make mesh & check resolution
mesh.pars <- c(max = 0.5,min = 0.09,cutoff = 0.2 )
mesh <- make.mesh(spatial.polygon = spatial.polygond, mesh.pars = mesh.pars)
plot.mesh(mesh) # Figure 5.1
### 6. Fit a joint model:  model fit this is; resp.1 = beta0 + field.1
    + made.up.cov
#resp.2 = alpha0 + field.2 + beta*field.1
```

```
fit <- joint.fit(mesh = mesh, locs.1 = locs.1, locs.2 = locs.2,
                response.1 = response.1, response.2 = response.2,
                family = c("binomial","binomial"),
                fixed.covariates.1 = fixed.covariates.1,
                fixed.covariates.2 = fixed.covariates.2,
                nl.cov.1 = nl.cov.1, nl.cov.2 = nl.cov.2, verbose = TRUE)
#Split DIC value of the joint model into components(HSS and DHSS)
tapply(fit$dic$local.dic,fit$dic$family,sum)

## function extracts the three random fields of the fitted model as a list
fields <- find.fields(fit, mesh = mesh, spatial.polygon = spatial.polygonc)

## to plot to correct spatial extent (Figure 5.6)
proj <- inla.mesh.projector(mesh)
## now extract the unique spatial estimated random field each variable
par(mfrow=c(1,2))
estRF.1 <- image.plot(proj$x,proj$y,fields[[1]],main="(a)",xlab="",ylab="")
plot(spatial.polygond,add=TRUE)
estRF.2 <- image.plot(proj$x,proj$y, fields[[2]],main="(b)",xlab="",ylab="")
plot(spatial.polygonc,add=TRUE)

## now extract the spatial estimated response for each variable,
just the inverse logit of the corresponding fixed effect
# (intercept) and the random field/fields (see model above)
# Maps of estimated HIV prevalence (Figure 5.4)
par(mfrow=c(1,2))
estResponse.1 <- image.plot(proj$x,proj$y,binomial(link="logit")$linkinv(
    fit$summary.fix[1,1] + fields[[1]]),main="(a)",xlab="",ylab="")
plot(spatial.polygond,add=TRUE)
estResponse.2 <- image.plot(proj$x,proj$y,binomial(link="logit")$linkinv(
    fit$summary.fix[2,1] + fields[[2]] + fit$summary.hyperpar[5,1]*fields
    [[1]]),main="(b)",xlab="",ylab="")
plot(spatial.polygonc,add=TRUE)
```

```
###7.Fit individual models and plot corresponding random effects
(no covariates in models):
fit.hss <- geo.fit(mesh = mesh,  locs = locs.1, response = response.1,
family = "binomial", verbose = TRUE)
fit.dhs <- geo.fit(mesh = mesh,  locs = locs.2, response = response.2,
family = "binomial", verbose = TRUE)
par(mfrow=c(1,2))
field.hss <- find.fields(fit.hss, mesh = mesh, spatial.polygon = spatial.polygond,
plot=TRUE)
estRF.1 <- image.plot(proj$x,proj$y,field.hss[[1]],main="(a)",xlab="",ylab="",
axes=FALSE)
plot(spatial.polygond,add=TRUE)

field.dhs <- find.fields(fit.dhs, mesh = mesh, spatial.polygon = spatial.polygonc)
estRF.2 <- image.plot(proj$x,proj$y,field.dhs[[1]],main="(b)",xlab="",ylab="",
axes=FALSE)
plot(spatial.polygonc,add=TRUE)
## ##now extract the spatial estimated response for each variable,
just the inverse logit of the corresponding fixed effect
# (intercept) and the random field/fields (see model above)

#Maps of estimated HIV prevalence (Figure 5.3)
par(mfrow=c(1,2))
estResponse.1 <- image.plot(proj$x,proj$y,binomial(link="logit")$linkinv(
fit.hss$summary.fix[1,1] + field.hss[[1]]),main="(a)",xlab="",ylab="")
plot(spatial.polygond,add=TRUE)
estResponse.2 <- image.plot(proj$x,proj$y,binomial(link="logit")$linkinv(
fit.dhs$summary.fix[1,1] + field.dhs[[1]]),main="(b)",xlab="",ylab="")
plot(spatial.polygonc,add=TRUE)

####8. Fit individual models with covariates
fit.hss <- geo.fit(mesh = mesh,  locs = locs.1, response = response.1,
family = "binomial", covariates = fixed.covariates.1 , nl.cov = nl.cov.1,
 verbose = TRUE)
fit.dhs <- geo.fit(mesh = mesh,  locs = locs.2, response = response.2,
family = "binomial", covariates = fixed.covariates.2 , nl.cov = nl.cov.2,
 verbose = TRUE)

### Plot spatial random effects (Figure 5.7)
field.hss <- find.fields(fit.hss, mesh = mesh, spatial.polygon =
spatial.polygon,plot=TRUE)
field.dhs <- find.fields(fit.dhs, mesh = mesh, spatial.polygon =
 spatial.polygon)
proj <- inla.mesh.projector(mesh)
par(mfrow=c(1,2))
estRF.1 <- image.plot(proj$x,proj$y,field.hss[[1]],main="(a)",
xlab="",ylab="",
        axes=FALSE)
plot(spatial.polygond,add=TRUE)
estRF.2 <- image.plot(proj$x,proj$y,field.dhs[[1]],main="(b)",
xlab="",ylab="",
        axes=FALSE)
plot(spatial.polygon,add=TRUE)



### 9. Plot of non-linear effects (Figure 5.5)

### Load shapefile
shape1 = readShapePoly("G:\\DHS2013 datasets\\districtric_boundaries
\\BDR_health districts")

shape1@data

### Create neighbourhood/adjacency matrix
# neig: neighbourhood structure
```

```
neig1 <- poly2nb(shape1)

# plot neighbourood
plot(neig1, coordinates(shape1), add=TRUE)

### Set up INLA call
# Convert it into an inla neighbourhood object
neig.inla = nb2INLA("neig1", neig1) # empty variable, file is saved

data1<-read.csv("G:\\DHS2013 datasets\\HSS_2014.csv")
#age as cont rand using random walk+BYM
forma=status˜-1+f(AgeDist,model="rw2")+gravida+f(Dis_ID, model="bym",
 graph="neig1")
resa1<-inla(forma,data=data1, family="binomial",control.compute=
list(dic=TRUE,cpo=TRUE)
,verbose = TRUE)
x.age1 <- resa1$summary.random$AgeDist$ID #not running
fhat.age1 <- resa1$summary.random$AgeDist$mean
max<-max(fhat.age)
min<-min(fhat.age)

plot(x.age1, fhat.age1, type="l",xlim=c(10,65),ylim=c(-1.6,1)
,main="(a)",xlab="Age in years",ylab="Effect")
lines(x.age1, resa1$summary.random$AgeDist$"0.025quant",lty=2,col="red")
lines(x.age1, resa1$summary.random$AgeDist$"0.975quant",lty=2,col="red")

###load data at constituency level##

shape2 = readShapePoly("G:\\Constituencies_shapefiles
\\Old_Constituency_Boundaries_107.shp")
shape2@data

### Create neighbourhood/adjacency matrix
```

```
# neig: neighbourhood structure
neig2 <- poly2nb(shape2)

# plot neighbourood
plot(shape2, border=gray(.5))
plot(neig2, coordinates(shape2), add=TRUE)

### Set up INLA call
# Convert it into an inla neighbourhood object
neig.inla = nb2INLA("neig2", neig2) # empty variable, file is saved

data2<-read.csv("G:\\DHS2013 datasets\\MEN_WOMEN_HIV_reduced.csv")
data2<-data.frame(data2)

#age a continuous var modelled using a random walk model

forma=HIV03~-1+f(age,model="rw2")+gender+kidsdead1+kidsdead2
+edu1+edu2+edu3+sexhhh+WealthIndexP+WealthIndexM+WealthIndexR+
WealthIndexRt+mstatus1+mstatus2+mstatus3+mstatus4+mstatus5+
agesex1+agesex2 +agesex3+agesex4+sexualactivity1+sexualactivity2
+condomUse
+STI+placeres+f(ID, model="bym", graph="neig2")
resa<-inla(forma,data=data2, family="binomial",control.compute=
list(dic=TRUE,cpo=TRUE))

#Ploting the random effects of age against cont.var:"Age"
x.age <- resa$summary.random$age$ID
fhat.age <- resa$summary.random$age$mean
max<-max(fhat.age)
min<-min(fhat.age)
par(mfrow=c(1,2))
plot(x.age1, fhat.age1, type="l",xlim=c(10,65),ylim=c(-1.6,1)
,main="(a)",xlab="Age in years",ylab="Effect")
```

```
lines(x.age1, resa1$summary.random$AgeDist$"0.025quant",lty=2,col="red")
lines(x.age1, resa1$summary.random$AgeDist$"0.975quant",lty=2,col="red")
plot(x.age, fhat.age, type="l",xlim=c(10,65),ylim=c(-3,1),
main="(b)",xlab="Age in years",ylab="Effect")
lines(x.age, resa$summary.random$age$"0.025quant",lty=2,col="red")
lines(x.age, resa$summary.random$age$"0.975quant",lty=2,col="red")
```

References

Adika, V., Baralate, S., Agada, J., & Nneoma, N. (2013). Mothers perceived cause and health seeking behaviour of childhood measles in Bayelsa Nigeria. *Journal of Research in Nursing and Midwifery*, *2*(1), 6–12.

Amornkul, P. N., Vandenhoudt, H., Nasokho, P., Odhiambo, F., Mwaengo, D., Hightower, A., … others (2009). HIV prevalence and associated risk factors among individuals aged 13-34 years in Rural Western Kenya. *PloS one*, *4*(7), e6470.

Arab, A. (2015). Spatial and spatio-temporal models for modeling epidemiological data with excess zeros. *International Journal of Environmental Research and Public Health*, *12*(9), 10536–10548.

Araújo, M. B., Thuiller, W., Williams, P. H., & Reginster, I. (2005). Downscaling european species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, *14*(1), 17–30.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data.* Boca Raton: Chapman & Hall.

Bao, L., Raftery, A. E., & Reddy, A. (2015). Estimating the sizes of populations at risk of HIV infection from multiple data sources using a Bayesian hierarchical model. *Statistics and its Interface*, *8*(2), 125.

Barankanira, E., Molinari, N., Niyongabo, T., & Laurent, C. (2016). Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention. *BMC Public Health*, *16*(1), 1.

Bellier, E., Neubauer, P., Monestiez, P., Letourneur, Y., Ledireach, L., Bonhomme, P., & Bachet, F. (2013). Marine reserve spillover: Modelling from multiple data sources. *Ecological Informatics*, *18*, 188–193.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, *14*(21-22), 2433–2443.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.

Besag, J., & Green, P. J. (1993). Spatial statistics and bayesian computation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25–37.

Besag, J., York, J., & Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *BF00116466*, *43*, 1–50.

Beyene, B. B., Tegegne, A. A., Wayessa, D. J., & Enqueselassie, F. (2016). National measles surveillance data analysis, 2005 to 2009, Ethiopia. *Journal of Public Health and Epidemiology*, *8*(3), 27–37.

Bhella, D., Bourhis, J., Combe, C., Cortay, J., Gahnnam, A., Gerlier, D., … Vidalain, P. (2007). *Measles virus Nucleoprotein.* New York: Nova Science.

Bisbe, J., Coenders, G., Saris, W. E., & Batista-Foguet, J. M. (2006). Correcting measurement error bias in interaction models with small samples. *Metodoloski Zvezki*, *3*(2), 267.

Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2008). *Applied spatial data analysis with R.* New York: Springer.

Blackwell, M., Honaker, J., & King, G. (2017). A unified approach to measurement error and missing data: Overview and applications. *Sociological Methods and Research*, *46*(3), 303–341.

Blangiardo, M., & Cameletti, M. (2015). *Spatial and Spatio-temporal Models with R-INLA.* UK: John Wiley & Sons.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and patio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, *7*, 39–55.

Bolstad, W. M. (2004). *Introduction to bayesian statistics.* New Jersey: John Wiley &Sons.

Buonaccorsi, P. J. (2010). *Measurement error models, methods, and applications.* Boca Raton: Chapmann & Hall.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models. A modern perspective.* Boca Raton: Chapmann & Hall.

Castro, H., Pillay, D., Sabin, C., & Dunn, D. T. (2012). Effect of misclassification of antiretroviral treatment status on the prevalence of transmitted hiv-1 drug resistance. *BMC Medical Research Methodology*, *12*(1), 30.

Charles, F. (2005). Strategies for dealing with measurement error in multiple regression. *Journal of Academy of Business and Economics*, *5*(3), 80.

Chege, W., Pals, S. L., McLellan-Lemal, E., Shinde, S., Nyambura, M., Otieno, F. O., ... Thomas, T. (2012). Baseline findings of an HIV incidence cohort study to prepare for future HIV prevention clinical trials in Kisumu, Kenya. *The Journal of Infection in Developing Countries*, *6*(12), 870–880.

Chen, X., Hong, H., & Nekipelov, D. (2007). Measurement error models. *Prepared for the Journal of Economic Literature. Retrieved from www. stanford. edu/˜ doubleh/eco273B/surveyjan27chenhandenis-07. pdf*.

Chiogna, M., & Gaetan, C. (2004). Hierarchical space-time modelling of epidemic dynamics: an application to measles outbreaks. *Statistical Methods & Applications*, *13*(1), 55–71.

Chipeta, M. G., Terlouw, D. J., Phiri, K., & Diggle, P. J. (2015). Adaptive geostatistical design and analysis for sequential prevalence surveys. *arXiv preprint arXiv:1509.04448*.

Congdon, P. D. (2010). *Applied Bayesian hierarchical methods.* Boca Raton: Chapmann & Hall.

Cowles, K. M., Yan, J., & Smith, B. (2009). Reparameterised and marginalised posterior and predictive sampling for complex Bayesian geostatistical models. *Journal of Computational and Graphical Statistics*, *18*, 262-282.

Cressie, N. A. C. (1993). *Statistics for spatial data.* New York: John Wiley & Sons.

Cunningham, R. B., & Lindenmayer, D. B. (2005). Modeling count data of rare species: Some statistical issues. *Ecology*, *86*(5), 1135–1142.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data.* New York: Chapman &Hall.

DeGroot, F. J. (1970). *Optimal statistical decisions.* New York: McGraw-Hill.

De Savigny, D., Mayombana, C., Mwageni, E., Masanja, H., Minhaj, A., Mkilindi, Y., ... Reid, G. (2004). Care-seeking patterns for fatal malaria in Tanzania. *Malaria Journal*, *3*(1), 27.

Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, *1186*(1), 125–145.

Diggle, P. J., & Ribeiro Jr, P. J. (2007). *Model-based geostatistics.* New York: Springer.

Diggle, P. J., Tawn, J., & Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(3), 299–350.

Dobbie, M. J., & Welsh, A. H. (2001). Models for zero-inflated count data using the Neyman type A distribution. *Statistical Modelling*, *1*(1), 65–80.

Dobson, A. J. (2002). *An introduction to generalised linear models.* New York: Chapman & Hall.

Doungmo, G. E. F., Oukouomi, N. S. C., & Mugisha, S. (2014). A fractional SEIR epidemic model for spatial and temporal spread of measles in metapopulations. In *Abstract and applied analysis* (Vol. 2014).

Downing, A., Forman, D., Gilthorpe, M. S., Edwards, K. L., & Manda, S. O. (2008). Joint disease mapping using six cancers in the Yorkshire region of England. *International Journal of Health Geographics*, *7*(1), 1.

Dwyer-Lindgren, L., Kakungu, F., Hangoma, P., Ng, M., Wang, H., Flaxman, A. D., … Gakidou, E. (2014). Estimation of district-level under-5 mortality in Zambia using birth history data, 1980–2010. *Spatial and spatio-temporal epidemiology*, *11*, 89–107.

Fernandes, M. V., Schmidt, A. M., & Migon, H. S. (2009). Modelling zero-inflated spatio-temporal processes. *Statistical Modelling*, *9*(1), 3–25.

Filia, A., Bella, A., Rota, M., Tavilla, A., Magurano, F., Baggieri, M., … Declich, S. (2013). Analysis of national measles surveillance data in italy from october 2010 to december 2011 and priorities for reaching the 2015 measles elimination goal. *Korea*, *1*, 4–5.

Finley, A. O., Banerjee, S., & Cook, B. D. (2014). Bayesian hierarchical models for spatially misaligned data in R. *Methods in Ecology and Evolution*, *5*(6), 514–523.

Fischer, M. M., & Getis, A. (2010). *Handbook of applied spatial analysis.* London: Springer.

Freedman, L. S., Fainberg, V., Kipnis, V., Midthune, D., & Carroll, R. J. (2004). A new method for dealing with measurement error in explanatory variables of regression models. *Biometrics*, *60*(1), 172–181.

Fuller, W. A. (1987). *Measurement error models.* New York: John Wiley & Sons.

Gelfand, A. E., Diggle, P. J., Fuentes, M., & Guttorp, P. (2010). *Hanbook of Spatial Statistics.* Boca Raton: Chapman & CRC/ Press.

Gelfand, A. E., & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, *4*(1), 11–15.

Gill, J. (2001). *Generalised linear models: A unified approach.* California: SAGE.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, *97*(458), 590–600.

Goovaerts, P. (2008). Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences*, *40*(1), 101–128.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632–648.

Government. (2002). *Education for all(EFA). National plan of action 2002-2015.* (Retrieved February 15, 2016 from http:// www. planipolis.iiep.unesco.org/sites/planipolis/files/.../namibia.efa.npa.pdf)

Government. (2017). *Namibia's 5th national development plan.* (Retrieved February 28, 2018 from http:// http://cirrus.com.na/namibias-fifth-national-development-plan/)

Griffith, D. A. (1985). An evaluation of correction techniques for boundary effects in spatial statistical analysis: contemporary methods. *Geographical Analysis*, *17*(1), 81–88.

Gschlößl, S., & Czado, C. (2008). Modelling count data with overdispersion and spatial effects. *Statistical Papers*, *49*(3), 531–552.

Guo, X., & Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, *58*(1), 16–24.

Gustafson, P. (2004). *Measurement error and missclassification in statistics and epidemiology: impacts and bayesian adjustments*. New York: Chapmann & Hall.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, *56*, 1030–1039.

Hampton, K. H., Serre, M. L., Gesink, D. C., Pilcher, C. D., & Miller, W. C. (2011). Adjusting for sampling variability in sparse data: Geostatistical approaches to disease mapping. *International Journal of Health Geographics*, *10*(1), 54.

He, Y., Landrum, M. B., & Zaslavsky, A. M. (2014). Combining information from two data sources with misreporting and incompleteness to assess hospice-use among cancer patients: A multiple imputation approach. *Statistics in Medicine*, *33*(21), 3710–3724.

Held, L., Höhle, M., & Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, *5*(3), 187–199.

Heymann, D. (2015). *Control of communicable diseases.* Washington DC: Apha Press.

Huo, X.-N., Zhang, W.-W., Sun, D.-F., Li, H., Zhou, L.-D., & Li, B.-G. (2011). Spatial pattern analysis of heavy metals in Beijing agricultural soils based on spatial autocorrelation statistics. *International Journal of Environmental Research and Public Health*, *8*(6), 2074–2089.

Huque, M. H., Bondell, H. D., Carroll, R. J., & Ryan, L. M. (2016). Spatial regression with covariate measurement error: A semiparametric approach. *Biometrics*, *72*(3), 678–686.

Illian, J. B., Møller, J., & Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, *16*(3), 389–405.

Isaaks, E. H., & Srivastava, M. R. (1989). *An introduction to applied geostatistics.* New York: Oxford University Press.

Jasem, J., Marof, K., Nawar, A., & Islam, K. M. (2012). Epidemiological analysis of measles and evaluation of measles surveillance system performance in Iraq, 2005–2010. *International Journal of Infectious Diseases*, *16*(3), e166–e171.

Jin, X., Carlin, B. P., & Banerjee, S. (2005). Generalized hierarchical multivariate car models for areal data. *Biometrics*, *61*(4), 950–961.

Kabaghe, A. N., Chipeta, M. G., McCann, R. S., Phiri, K. S., Van Vugt, M., Takken, W., … Terlouw, A. D. (2017). Adaptive geostatistical sampling enables efficient identification of malaria hotspots in repeated cross-sectional surveys in rural Malawi. *PloS one*, *12*(2), e0172266.

Kamdem, C., Fouet, C., Etouna, J., Etoa, F.-X., Simard, F., Besansky, N. J., & Costantini, C. (2012). Spatially explicit analyses of anopheline mosquitoes indoor resting density: implications for malaria control. *PloS one*, *7*(2), e31843.

Keil, P., Belmaker, J., Wilson, A. M., Unitt, P., & Jetz, W. (2013). Downscaling of species distribution models: A hierarchical approach. *Methods in Ecology and Evolution*, *4*(1), 82–94.

Keller, G. (2012). *Managerial statistics.* Ohio: Mason.

Knorr-Held, L. (1999). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, *19*(17-19), 2555-2567.

Knorr-Held, L., & Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(1), 73–85.

Krainski, E., & Lindgren, F. (2013). *The R-INLA tutorial: SPDE models.* Technical report, Available online at www. r-inla. org.

Kroese, D. P., Taimre, T., & Zadravko, B. I. (2011). *Handbook of Monte Carlo methods.* New Jersey: John Wiley & Sons.

Kumar, V., Chaudhury, L., Rathore, R., Taneja, D., Ramnath, R., & Bhushan, B. (2003). An epidemiological analysis of outbreak of measles in a medical relief camp. *Population and Health Perspectives and Issue*, *26*(4), 135–140.

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

Lawson, A. B., Biggeri, A., Bohning, D., Lesaffre, E., Viel, J.-F., & Bertollini, R. (1999). *Disease mapping and risk assessment for publick health.* New York: John Wiley & Sons.

Lawson, A. B., & Williams, F. L. R. (2013). *An introductory guide to disease mapping.* UK: John Wiley & Sons.

Lee, D., & Sarran, C. (2015). Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. *Environmetrics*, *26*(7), 477–487.

Lentz, J. A., Blackburn, J. K., & Curtis, A. J. (2011). Evaluating patterns of a white-band disease (WBD) outbreak in Acropora palmata using spatial analysis: a comparison of transect and colony clustering. *PloS one*, *6*(7), e21830.

Li, R., Conti, D. V., Diaz-Sanchez, D., Gilliland, F., & Thomas, D. C. (2013). Joint analysis for integrating two related studies of different data types and different study designs using hierarchical modeling approaches. *Human Heredity*, *74*(2), 83–96.

Liang, S., Banerjee, S., Bushhouse, S., Finley, A. O., & Carlin, B. P. (2008). Hierarchical multiresolution approaches for dense point-level breast cancer treatment data. *Computational Statistics & Data Analysis*, *52*(5), 2650–2668.

Lindgren, F. (2012). Continuous domain spatial models R-INLA. *ISBA Bulletin*, *19*(4), 423–498.

Lindgren, F., & Rue, H. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. *Journal of Royal Statistical Society*, *73*, 423–498.

Lindsey, J. K. (2001). *Applying generalised linear models: A unified approach.* New York: Springer.

Ma, C., Hao, L., Zhang, Y., Su, Q., Rodewald, L., An, Z., ... others (2014). Monitoring progress towards the elimination of measles in china: an analysis of measles surveillance data. *Bulletin of the World Health Organization*, *92*(5), 340–347.

Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science & Medicine*, *55*(1), 125–139.

Manda, S., Feltbower, R., & Gilthorpe, M. (2012). Review and empirical comparison of joint mapping of multiple diseases. *Southern African Journal of Epidemiology and Infection*, *27*(4), 169–182.

Manda, S., Masenyetse, L., Cai, B., & Meyer, R. (2015). Mapping HIV prevalence using population and antenatal sentinel-based HIV surveys: a multi-stage approach. *Population Health Metrics*, *13*(1), 1.

Mayet, A., Genicon, C., Duron, S., Haus-Cheymol, R., Ficko, C., Bédubourg, G., ... others (2013). The measles outbreak in the french military forces–2010–2011: results of epidemiological surveillance. *Journal of Infection*, *66*(3), 271–277.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models.* New York: Chapman & Hall.

Mohebbi, M., Wolfe, R., & Forbes, A. (2014). Disease mapping and regression with count data in the presence of overdispersion and spatial autocorrelation: A Bayesian model averaging approach. *International Journal of Environmental Research and Public Health*, *11*(1), 883–902.

MoHSS. (2013). *Namibia demographic and health survey 2013.* (Retrieved April 15, 2016 from http:// www.dhsprogram.com/pubs/pdf/FR298/FR298.pdf)

MoHSS. (2014a). *Namibia child survival strategy 2014-2018.* (Retrieved March 20, 2016 from https://www.medbox.org/namibia/namibia-child-survival-strategy-2014-2018 /preview?...)

MoHSS. (2014b). *Surveillance report of 2014 National HIV sentinel survey* (Tech. Rep.). (Retrieved April 15, 2016 from http:// www.mhss.gov.na/.../12f-2014-National-HIV-Sentinel-Survey .pdf)

Mufune, P., Kaundjua, M. B., & Kauari, L. (2014). Young peoples perceptions of sex and relationships in northern Namibia. *International Journal of Child, Youth and Family Studies*, *5*(2), 279–295.

Musenge, E., Vounatsou, P., Collinson, M., Tollman, S., & Kahn, K. (2013). The contribution of spatial analysis to understanding HIV/TB mortality in children: a structural equation modelling approach. *Glob Health Action*, *6*, 38–48.

Naish, S. (2012). *A spatial temporal analysis of Barmah forest virus disease in Queensland, Australia* (Unpublished doctoral dissertation). School of Public Health, Queensland University of Technology, Australia.

Neyens, T., Lawson, A. B., Kirby, R. S., Nuyts, V., Watjou, K., Aregay, M., ... Faes, C. (2017). Disease mapping of zero-excessive mesothelioma data in flanders. *Annals of epidemiology*, *27*(1), 59–66.

Ngesa, O. (2014). *Bayesian spatial models with application to HIV, TB and STI in Kenya* (PhD dissertation). College of Agriculture, Engineering and Science. School of Mathematics, Statistics and computer Science. University of KwaZulu-Natal, South Africa.

Ngesa, O., Mwambi, H., & Achia, T. (2014). Bayesian spatial semi-parametric modeling of HIV variation in Kenya. *PloS one*, *9*(7), e103299.

Njai, R., Siegel, P. Z., Miller, J. W., & Liao, Y. (2011). Misclassification of survey responses and black-white disparity in mammography use, behavioral risk factor surveillance system, 1995-2006. *Prev Chronic Dis*, *8*(3), A59.

NPC. (2017). *Millennium development goals.* (Retrieved February 28, 2018 from http:// www.un.org/millenniumgoals/.../pdf/MDG. pdf)

Ntirampeba, D., Neema, I., & Kazembe, L. (2017). Modelling spatial patterns of misaligned disease data: An application on measles incidence in Namibia. *Clin Epidemiol Glob Health*.

Ntzoufras, I. (2009). *Bayesian modelling using WinBUGS*. New York: John Wiley &Sons.

Okango, E., Mwambi, H., Ngesa, O., & Achia, T. (2015). Semi-Parametric Spatial Joint Modeling of HIV and HSV-2 among Women in Kenya. *PloS one*, *10*(8), e0135212.

Onicescu, G., Hill, E. G., Lawson, A. B., Korte, J. E., & Gillespie, M. B. (2010). Joint disease mapping of cervical and male oropharyngeal cancer incidence in blacks and whites in South Carolina. *Spatial and Spatio-temporal Epidemiology*, *1*(2), 133–141.

Paireau, J., Girond, F., Collard, J.-M., Maïnassara, H. B., & Jusot, J.-F. (2012). Analysing spatio-temporal clustering of meningococcal meningitis outbreaks in niger reveals opportunities for improved disease control. *PLoS Negl Trop Dis*, *6*(3), e1577.

Pan, W., Jeong, K. S., Xie, Y., & Khodursky, A. (2008). A nonparametric empirical Bayes approach to joint modeling of multiple sources of genomic data. *Statistica Sinica*, 709–729.

Peuquet, D. J. (1999). Making space for time: Issues in space-time data representation. In *Database and expert systems applications, 1999. proceedings. tenth international workshop on* (pp. 404–408).

Pfeiffer, D. U., Robinson, T. P., Stevenson, M., Stevens, K. B., Rogers, D. J., & Clements, A. C. (2008). *Spatial analysis in epidemiology.* New York: Oxford University Press.

Pischke, S. (2007). Lecture notes on measurement error. *London School of Economics, London.*

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Raghunathan, T. E., Xie, D., Schenker, N., Parsons, V. L., Davis, W. W., Dodd, K. W., & Feuer, E. J. (2007). Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening. *Journal of the American Statistical Association*, *102*(478), 474–486.

Ratkowsky, D. A. (1990). *Nonlinear regression models for repeated measurement data.* New York: Marcel Dekker.

Restrepo, A. C., Baker, P., & Clements, A. C. (2014). National spatial and temporal patterns of notified dengue cases, Colombia 2007–2010. *Tropical Medicine and International Health*, *19*(7), 863–871.

Ridout, M., Demétrio, C. G., & Hinde, J. (1998). Models for count data with many zeros. In *Proceedings of the 19th international biometric conference* (Vol. 19, pp. 179–192).

Riebler, A., Srbye, S., Simpson, D., & Rue, H. (2016). *An intuitive Bayesian spatial model for disease mapping that accounts for scaling.* arXiv.

Robert, C., & Casella, G. (2010). *Introducing Monte Carlo methods with R.* New York: Springer.

Robert, C., & Casella, G. (2011). A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, *26*(1), 102–115.

Rodriques-Motta, M., Gianola, D., Heringstad, B., Rosa, G. J., & Chang, Y. (2007). A zero-inflated poisson model for genetic analysis of the number of mastitis cases in norwegian red cows. *Journal of Dairy Science*, *90*(11), 5306–5315.

Roli, G., & Raggi, M. (2015). Bayesian hierarchical models for misaligned data: a simulation study. *Statistica*, *75*(1).

Roussas, G. G. (1997). *A course in mathematical statistics (2nd ed.).* New York: Academic Press.

Samaniego, F. J. (2010). *A comparison of bayesian and frequentist approaches to estimation.* New York: Springer.

Schenker, N. (2013). *Combining information from multiple data systems to enhance analyses related to health: Examples and lessons learned.* (Retrieved from http://www.semanticscholar.org/.../Combining-Information-from-Multiple-Data-System/)

Schenker, N., Gentleman, J. F., Rose, D., Hing, E., & Shimizu, I. M. (2002). Combining estimates from complementary surveys: A case study using prevalence estimates from national health surveys of households and nursing homes. *Public Health Reports*, *117*(4), 393–407.

Schenker, N., Raghunathan, T. E., & Bondarenko, I. (2010). Improving on analyses of self-reported data in a large-scale health survey by using information from an examination-based survey. *Statistics in Medicine*, *29*(5), 533–545.

Scott, L. M. (2007). *Introduction to applied bayesian statistics and estimation for social scientists.* New York: Springer.

Sherman, M. (2011). *Spatial statistics and spatio-temporal data.* UK: John Wiley & Sons.

Smith, B. J., Yan, J., Cowles, M. K., et al. (2008). Unified geostatistical modeling for data fusion and spatial heteroskedasticity with R package ramps. *Journal of Statistical Software*, *25*(10), 1–21.

Song, H.-R., Lawson, A., Agostino, R. B., & Liese, A. D. (2011). Modeling type 1 and type 2 diabetes mellitus incidence in youth: An application of Bayesian hierarchical regression for sparse small area data. *Spatial and Spatio-temporal Epidemiology*, *2*(1), 23–33.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modelling for a complex world.* New York: Irwin & Mac Grau-hill.

Sturrock, H. J., Cohen, J. M., Keil, P., Tatem, A. J., Le Menach, A., Ntshalintshali, N. E., ... Roland D, G. (2014). Fine-scale malaria risk mapping from routine aggregated case data. *Malaria Journal*, *13*(1), 421.

Sturrock, H. J., Pullan, R. L., Kihara, J. H., Mwandawiro, C., & Brooker, S. J. (2013). The use of bivariate spatial modeling of questionnaire and parasitology data to predict the distribution of Schistosoma haematobium in coastal Kenya. *PLoS Negl Trop Dis*, *7*(1), e2016.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*(4), 1701–1728.

UNAIDS. (2013). *Efficient and sustainable HIV responses: Case studies on country progress.* (Retrieved October 20, 2015 from http://www.unaids.org/.../20130115-JC2450-case-studies-country-progress)

UNAIDS. (2015a). *Fact sheet 2015* (Tech. Rep.). (Retrieved March 15, 2016 from http:// www.unaids.org/sites/default/.../20150901.FactSheet.2015.en.pdf)

UNAIDS. (2015b). *Global AIDS response progress reporting 2015.* (Retrieved February 15, 2016 from http:// www.unaids.org/sites/default/files.pdf)

Van Beers, W. C., & Kleijnen, J. P. C. (2003). Kriging for interpolation in random simulation. *Journal of the Operational Research Society*, *54*(3), 255–262.

Wackerly, D. D., Mendenhall III, W. L., & Scheaffer, R. L. (2002). *Mathematical statistics with applications (6th ed.).* New York: Duxbury.

Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data.* New Jersey: John Wiley & Sons.

Wattanasaruch, P., Pongsapukdee, V., & Khawsithiwong, P. (2012). Least-MSE calibration procedures for corrections of measurement and misclassification errors in generalized linear models. *Songklanakarin Journal of Science and Technology*, *34*(4).

WHO. (2014). *WHO warns that progress towards eliminating measles has stalled.* (Retrieved February 15, 2016 from http://www.who.int/mediacentere/news/release/2014/eliminating-measles/en/)

WHO. (2015). *Measles.* (Retrieved October 20, 2015 from http://www.who.int/mediacentre/factsheets/fs286/en/)

WHO. (2017). *Reported measles cases and incidence rates by who member states 2013, 2014 as of 11february 2015.* (Retrieved October 16, 2016 from http://www.who.int/immunization/.../measlesreportedcasesbycountry.pdf)

Xia, H., & Carlin, B. P. (1998). Spatio-temporal models with errors in covariates: mapping ohio lung cancer mortality. *Statistics in Medicine*, *17*(18), 2025–2043.

Yamada, I., & Rogerson, P. A. (2003). An empirical comparison of edge effect correction methods applied to k-function analysis. *Geographical Analysis*, *35*(2), 97–109.

Yan, J., Cowles, M. K., Wang, S., & Armonstrong, M. P. (2007). Parallelising MCMC for Bayesian spatiotemporal geostatistical models. *Statistics and Computing*, *17*, 323–335.

Yi, L., Tang, H., & Lin, X. (2009). Spatial linear mixed models with covariate measurement errors. *Statistica Sinica*, *19*(3), 1077.

Zagheni, E., Billari, F. C., Manfredi, P., Melegaro, A., Mossong, J., & John W, E. (2008). Using time-use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*, *168*(9), 1082–1090.

Zhu, Y., Xu, Q., Lin, H., Yue, D., Song, L., Wang, C., … Li, X. (2013). Spatiotemporal analysis of infant measles using population attributable risk in shandong province, 1999–2008. *PloS one*, *8*(11), e79334.