BAYESIAN SPATIAL MODELLING OF *TUBERCULOSIS* AND ITS EFFECTS ON SOCIO-ECONOMIC AND DEMOGRAPHIC FACTORS IN SOUTH AFRICA: A CASE STUDY OF THE EASTERN CAPE PROVINCE.

University of Fort Hare

Together in Excellence

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY IN BIOSTATISTICS

BY

OBAROMI, ABIODUN DAVIES

(201407510)

DEPARTMENT OF STATISTICS

FACULTY OF SCIENCE AND AGRICULTURE

UNIVERSITY OF FORT HARE

ALICE, 5700

SUPERVISOR: PROFESSOR YONGSONG QIN

CO-SUPERVISOR: DR. JAMES NDEGE

2018

Declaration

I, **MR. OBAROMI ABIODUN DAVIES**, declare that I have prepared this thesis myself titled, **"BAYESIAN SPATIAL MODELLING OF TUBERCULOSIS AND ITS EFFECTS ON SOCIO-ECONOMIC AND DEMOGRAPHIC FACTORS IN SOUTH AFRICA: A CASE STUDY OF THE EASTERN CAPE PROVINCE."** and the work presented in it is my own. I confirm that:

This work was done exclusively and primarily while in candidature for a research doctoral degree at the University of Fort Hare, Alice.

No part or section of this thesis has previously been submitted as part of any application for a degree or any other qualification at this University or any other institution.

Where I have referred the published work of others, this is always clearly recognized.

Where I have cited from the work of others, the source is always given.

With the exception of such citations, this thesis is entirely my own work.

Name: Obaromi, A. D.

Signed:  Date: _19/03/2018_

Approved by Supervisor(s):

Supervisor: Prof. Yongsong Qin Signed:  Date: 23 /03/ 2018

Co-supervisor: Dr. James Ndege Signed:_____ Date: 23/ 03/ 2018

Dedication

This thesis is dedicated to God Almighty, the giver of life, my helper and my sustenance, and also to the memories of those who have passed on as a result of the scourge of *Tuberculosis*, especially my late elder brother, Mr. Ayodeji Olufemi Andrew Obaromi (1974-1999).

Acknowledgements

Firstly, I wish to greatly thank the almighty God for bringing me this far in my educational life and granting me His mercies and also for the gift of life. Special thanks go to my supervisor Professor Yongsong Qin and my co-supervisor, Dr. James Ndege, for their immense contributions right from inception of this doctoral programme. Their untiring and indepth supervision, alongside their constructive and thorough criticisms toward the emergence of a quality thesis will always be remembered. I owe this duo my continuous appreciation as I move up the ladder as a career biostatistician.

I would like to thank the management of Kogi State Polytechnic, Lokoja, Nigeria for granting me the study leave that made this work possible. Also to all my colleagues in the Department of Maths/Statistics for their support while I was away, in a particular manner, Mr. S. S. Ikusika, Mr. F. K. Lawal, Mr. T.J. Alabi, Mr. Peter Idoko, Mrs. Ayodele, Mr. Julius Omatola and other colleagues. I equally appreciate my sponsoring body, Tertiary Education Trust Fund (TETFUND) Abuja, Nigeria, for the scholarship. Also to all my friends that I met while studying in South Africa, most of who assisted during my data collection and other logistics; special thanks to Okango Elphas, Shirley Dali-Beza, Wendy Dali, Ms Babongile, Mr. Merile, Mrs. Nano, Ms Solani Ncences, Mrs. Ntomboxolo Masiza, Mrs. Faith, Mr. Madoba, Tunyiwe Honono and a good host of others.

Thanks to my mentors, colleagues and friends at the University of Fort Hare, Alice campus, for providing a pleasant and inspiring working environment. Thanks to Prof. and Prof. (Mrs) Anthony Okoh, Prof. Uche, Dr. F.K. Okaiyeto, Dr. Emmanuel Segun, Dr. Chris Baiyegunhi, Dr. Tony Anukam, Mr. Odeyemi, Dr. Ruffin Mutambayi, Dr. Zee, Dr. Goke Adeniji, Dr. AbdulrasaqYahaya,

Dr. Balogun Ayodeji, Roland, Azeez Ajiboye, e.t.c., and to all my wonderful family of the Redeemed Christian Church of God, Fountain of Life Parish, for their support and prayers. God bless you all real good!

Last but not the least, my biggest appreciation goes to my family for their continuous support and for keeping me relevant in their lives while I was away. I thank my wife, Olufunke Comfort Obaromi immensely and my children; Edmond, Edna, Edith and Edeline for their unfailing patience and prayers all through my stay here. You have always been my encouragement throughout this painstaking period of my educational life. I love you guys. I also thank my siblings for their love and encouragement in the course of this programme. In a special way I would like to thank my loving parents, Mr. and Mrs. Jerome kayode for their love, encouragement, prayers and for seeing me through my education. May you live long to reap the fruits of your untiring labour.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation is concerned with evolving and extending statistical models in the area of Bayesian spatial modelling, an increasingly important field of spatial epidemiology with particular interest towards application to Tuberculosis data in the Eastern Cape province of South Africa. In spatial epidemiology, the diseases to be examined usually occur within a map that needs spatial statistical methods that are appropriate, to model the observed data in the presence of some covariates and also cater for the variation of the disease.

In this thesis, the Bayesian models were developed in such a way that they allowed several factors classified as fixed and random effects, to be included in the models and using the Bayesian approach. The basic model used in disease mapping is the Besag, York and Mollie model, which incorporates two random effects; one which is spatially structured and the other random effect which is spatially unstructured. The effects (fixed and random) were the covariate effects, socio-economic and demographic variability and the spatial variability respectively, which were all investigated in seven different hierarchical/multilevel Bayesian models. These factors showed varying and substantial effects in the posterior relative risk estimation of the disease.

We assumed a negative binomial and generalized Poisson distributions to the response variable or relative risk estimate, y_i, to capture the over-dispersion phenomenon that is common and inherent with Poisson density for counts data. Spatial and non-spatial models were developed to model over-dispersion with all the distributions; Poisson, negative binomial and generalized Poisson. Negative binomial and generalized Poisson showed varying properties from comparisons with DIC values and parameter estimates to standard errors, which made either of them fit depending on the choice of model selection.

It was found that a lower DIC value could be insufficient to determine a best fit model, if other models present estimates with lower variances even at higher DIC values. The generalized Poisson, a two parameter distribution like the negative binomial, which also has the ability to capture both under-dispersion and over-dispersion, was found to perform better in the results than the negative binomial on the basis of a lower variance and with more exact parameter estimates.

A new weighted prior distribution, the "Besag2" ICAR model for the structured spatial random effects, which is an extension of the traditional ICAR prior model with two hyperparameters, was also developed and compared with some existing prior models; BYM and ICAR, to measure for spatial dependency in the regions. This new prior distribution was found to show a better fit, when compared to the basic ICAR prior usually assumed for the spatial random effect in the BYM model. This newly parameterized prior distribution in the Besag, York and Mollie model also led to improved parameter control, as the hyperparameters can be seen independently from each other. The result also showed that the new model performed well, both presenting good learning abilities and good shrinkage behaviour. In terms of model choice criteria, the proposed model performed at least equally well and better than the existing models, and the new formulation also gave parameters that are interpretable and have a clearer meaning.

To interpolate scattered or regularly distributed data, there are imprecise or exact methods, but there are some of these methods that could be used for interpolating data in a regular grid and others in an irregular grid. Linear and biharmonic spline methods were implemented in MATLAB, to compare for smoothing in the distribution patterns of tuberculosis in the province. This smoothing spline is a method of fitting a smooth curve to a set of noisy observations using a spline function.

This new method is rarely used in disease mapping applications, but it has a superior advantage to be assessed at subjective locations rather than only on a rectangular grid as seen in most traditional GIS methods of geospatial analyses.

The proposed new models and methods in this thesis were found to be flexible and robust, since they can be reduced or extended according to the nature of the data. Nevertheless, great care must be considered in the choice of prior densities. The approaches developed in this dissertation helped to broaden the scope for spatial analysis and disease mapping applications in epidemiology and public health studies.

Abbreviations

AIC	**A**kaike **I**nformation **C**riterion
AIDS	**A**cquired **I**mmune **D**eficiency **S**yndrome
ATS	**A**merican **T**horacic **S**ociety
BIC	**B**ayesian **I**nformation **C**riterion
BYM	**Besag, Y**ork and **M**ollie
CAR	**C**onditional **A**utoregressive
CDC	**C**enter for **D**isease **C**ontrol
CDS	**C**enter for **D**isease **S**tudy
CF	**C**haracteristic **F**unction
DF	**D**egree of **F**reedom
DIC	**D**eviance **I**nformation **C**riterion
ECAC	**E**astern **C**ape **A**IDS **C**ouncil
ECSECC	**E**astern **C**ape **S**ocio **E**conomic **C**onsultative **C**ouncil
GGD	**G**eneralized **G**aussian **D**istribution
GIS	**G**eographical **I**nformation **S**ystem
GLM	**G**eneralized **L**inear **M**odels
GLMM	**G**eneralized **L**inear **M**ixed **M**odels
GP	**G**eneralized **P**oisson
GRF	**G**aussian **R**andom **F**ield
HDI	**H**uman **D**evelopment **I**ndex
HIV	**H**uman **I**mmunodeficienc**y** **V**irus
ICAR	**I**ntrinsic **C**onditional **A**utoregressive

INLA	**I**ntegrated **N**ested **L**aplace **A**pproximation
IQU	**I**ndex of **Q**uality of **U**rban
LISA	**L**ocal **I**ndicators **of** **S**patial **A**ssociation
LRT	**L**ikelihood **R**ati**o** **T**est
MCMC	**M**arkov **C**hain **M**onte **C**arlo
MDR-TB	**M**ulti-**D**rug **R**esistance **T**uberculosis
MLE	**M**aximum **L**ikelihood **E**stimation
MRC	**M**edical **R**esearch **C**ouncil
MRF	**M**arkov **R**andom **F**ield
MTB	**M**ycobacterium **T**uberculosis
NBD	**N**egative **B**inomial **D**istribution
NDOH	**N**ational **D**epartment of **H**ealth
PTB	**P**ulmonary **T**uberculosis
SAR	**S**imultaneous **A**utoregresssive
SES	**S**ocio **E**conomic **S**tatus
SIR	**S**tandardised **I**ncidence **R**atio
SMR	**S**tandard **M**ortalit**y** **R**atio
TB	**T**uberculosis
WHO	**W**orld **H**ealth **O**rganisation
XDR-TB	**E**xtensive **D**rug **R**esistant **T**uberculosis
ZINB	**Z**ero **I**nflated **N**egative **B**inomial
ZIP	**Z**ero **I**nflated **P**oisson

Chapter One

General Introduction

1.1 Overview

A growing interest in the advance and use of spatial statistical methods in analysing geographically correlated data has increased in the recent times. This can be connected to the increasing availability and accessibility to data that are geo-referenced, especially in public health and ecology fields of study. Most data obtained from governments and agencies through surveys are demarcated geographically by quarters, regions, provinces or other administrative units. It is important to note that the population level risk of a particular disease can vary across geographical regions, and it is of great concern to governments, health authorities and policy makers to discover these variations in disease risk in order to identify possible underlying reasons for these differences. Most disease mapping methods are based on the geographical region being divided into areal units, with the disease risks been estimated for each of these areas. This is because individual level data would be against patient's confidentiality, and because governments are more interested in risk levels for populations as a whole. This thesis is inclined to multilevel models in disease mapping for count data. This is purely due to the accessibility and nature of such data in this situation. This thesis is also introduced by discussing key concepts in spatial modelling for a single disease model.

1.2 Disease Mapping

Disease mapping is usually carried out to investigate the geographical distribution of disease burden. Area specific estimates of risk may inform public health resource allocation by estimating the disease burden in specific areas, and the informal comparison of risk maps with exposure maps may provide clues to the etiology or generate hypotheses (Wakefield 2007). The applications of disease mapping in epidemiology and public health have enjoyed a great deal of acceptance. Disease mapping also refers to the estimation and demonstration of summary measures of health outcomes that are spatially observed (Rezaeian *et al.*, 2007). Some of the purposes of disease mapping include, to;

1. describe geographical variation and distribution of diseases.
2. generate hypotheses about a disease dynamics.
3. generate disease atlases.
4. detect clustering and characterization of a disease.

To introduce this method of study, it is required that one defines the geographic scale for data collection and analysis. From literature, data collected for disease mapping is generally categorized as either an areal data or point level data and both are described briefly here. At the point level, we consider those data collected at a set of sites or regions; say c, which are considered to vary continuously over the area under study. For example, point level data arises when we know the geographic locations of individual cases of disease, expressed in terms of latitude and longitude. For this reason, point level data may also be described as geocoded or geostatistical data (Banerjee *et al.*, 2004).

Given these data, it is often of interest to infer the distribution of the recorded outcome over the entire region of interest, based only on information collected on *c*, and taking spatial correlation into account. This spatial correlation is often measured as a function of distance between pairs of locations in *c*, with measurement aided by Geographic Information Systems (GIS) technology. While useful for capturing information on individual cases, the use of point data is not widespread in disease mapping, due to issues surrounding individual case confidentiality.

Areal data conversely, involves the definition of geographic boundaries, either regular or irregular, such that individual cases of disease are combined based on the regions formed. A common example of boundary specification is by Statistical Local Area (SLA) (ABS, 2006). Compared to point data, areal data is easily obtainable, such as data from census and other government agency databases.

Mapping of disease incidence and prevalence is a common concept in public health and epidemiology. Regularly, the main interest in disease mapping is to smoothen and predict some response variables over a geographical area of interest. However, there are two central characteristics of disease mapping, namely geographical distribution and the disease. The area-specific estimates of the diseases can be used by policy makers to prioritize and make decisions on public health resources allocation and interventions. Two sources of variability that often appear in disease mapping studies at the area level that violate statistical assumptions for the residual term of the model is the unexplained variation and which can be divided into two parts: a structured (spatially correlated) and an unstructured (spatial random) component.

The concept of disease mapping can also be used to describe geographical variation of diseases, identify clustering of diseases and generate atlas of diseases. A good number of statistical reviews on disease mapping have been done (Wakefield, 2007; Clayton and Bernardinelli, 1992;

Smans and Esteve, 1997; Wakefield *et al.,* 2000; Manda *et al.,* 2011). Disease mapping has a long history in epidemiology (Walter, 2000) as part of the classic triad of person/ place/time. A number of statistical reviews are available (see for example, Smans and Esteve, 1992; Clayton and Bernardinelli, 1992; Mollie, 1996; Wakefield *et al.,* 2000).

Disease mapping has seen many applications in epidemiology and public health. The backbone model for univariate disease mapping is the Besag, York and Mollie (BYM) model proposed by Besag *et al.,* (1991) and was also the first to incorporate spatial smoothing into studies of disease mapping in a fully Bayesian framework. This work was an extension of earlier work by Clayton and Kaldor (1987), who considered the use of Empirical Bayes for the estimation of spatially correlated relative risks on the logarithmic scale.

1.3 A Regression Model for Count Data

In some epidemiological or clinical studies, the response of interest consists of counts, such as the number of cells that show defective evidence of differentiation, or the number of repeated infections experienced by a subject. The values recorded are always non-negative integers. In some cases, it may be possible to analyse observed data that are counted by the methods of multiple linear regression. Nevertheless, regression models that are available and better suitable for response measurements that are counted such as disease counts, in an area say *i*, include Poisson regression, negative binomial regression and zero-inflated models (David *et al.,* 2007). Since it regularly provides a suitable representation for the variability observed in count data, the Poisson distribution plays a part in the analysis that is similar to that of the normal distribution in multiple linear regressions.

As generally known, that the frequently used model for count regression data is the Poisson regression. One of the key features of the Poisson distribution is that the variance equals the

mean, but empirically, however, we often find data that exhibit over-dispersion, with a variance larger than the mean (Rodriguez 2013). Therefore, the concept of over-dispersion is more widely applicable in count data for the Poisson regression model. Over-dispersion can be termed as an extra variation when the systematic structure of the model is correct.

Over-dispersion stems from how the stochastic component of the model is outlined. In a number of practical conditions where we desire to model count data, we may have variable observation periods for our counts (Richard *et al.,* 2007). Over-dispersion arises when the marginal variance of an attribute exceeds the theoretical variance implied by a chosen model (Robert *et al.,* 2008). The concepts and implications of a large variance in statistical models have always posed concerns. It is generally known that the larger the variance, the more imprecise our estimates and vice-versa. Also in practice, however, this assumption of equi-dispersion is often false, since the variance can either be larger or smaller than the mean, that is, both over-dispersion and under-dispersion can exist in count data.

Whenever we identify the possible presence of over-dispersion in count data, what could be the consequences of failing to take it into account? Firstly, the standard errors obtained from the model will be incorrect and unfitting and may be seriously underestimated and therefore, we may incorrectly assess the significance of individual regression parameters (John, 2007). Also, changes in deviance associated with model terms will also be too large and this will lead to the selection of overly complex models. Finally, the interpretation of the model will be incorrect and this may affect the prediction accuracy.

Statistically, if over-dispersion is present, disregarding it is likely to have little influence on point estimates of the regression parameters (the maximum likelihood estimator is consistent, although some small sample bias might be present).

Though, standard error estimates for regression parameters are underestimated. Type I errors related with testing whether regression coefficients are zero or not are underestimated, which is particularly problematic in relation to covariates that are close to the significance threshold. If the objective is to build a parsimonious model, the presence of over-dispersion may result in an analyst building a model more complex than necessary, and that overestimates the variance explained. These statistical problems are similar to those that are encountered when fitting models where there is spatial autocorrelation in model residuals (Haining 1990, 2003). When over-dispersion occurs, we modify the model and also keep the Poisson model, but add ad-hoc models for the variance. Specifying alternative distributions that can generate over-dispersion can also be employed.

For this thesis, we shall be discussing single disease (TB) modelling. Epidemiologic models like Bayesian regression models from the Besag, York and Mollie model (BYM) and geostatistics information are useful in this regard by allowing an assessment of the influence a disease outbreak may have under a selection of conditions. Such models are recognised as central to developing management strategies for disease.

This dissertation also tries to show the incidence and prevalence of tuberculosis based on some demographic, socioeconomic and transmission-related indicators as an important determinant of disparities in TB rates in the Eastern Cape Province of South Africa.

1.4 Statement of the Research Problem

Some models have been established for both single and multiple diseases at ecological levels to deal with their incidences and prevalence. Most of these models are based on the use of Bayesian regression models and some are connected with random effects, which are divided into spatial and non-spatial components.

The presence of over-dispersion is a particular problem for the analysis of spatially grouped data such as counts of events for census tracts that tend to have populations that socially, economically and demographically are heterogeneous or where the event has a tendency to cluster at the tract scale perhaps due to an infectious process. Over-dispersion, which may arise as a result of data imprecision or choosing an unsuitable probability model needs to be adjusted with suitable probability distributions applied to give a better data estimate when observed.

Also, violations of equi-dispersion from Poisson model show potential correlation in the data, which can affect the standard errors of parameter estimates. Therefore, alternative approaches to over-dispersion include reparameterisations of the variance function. This requires that alternative models, a type of mixture models built from Poisson be employed, to model and capture various types of dispersions in count data, which have predictable implications for the probabilistic structures of such models. Some of these mixture models also have the capacity to account for under-dispersion, though in very rare cases.

Another challenge in spatial analyses is the problem of spatial autocorrelation. Modelling the spatial interactions that arise in spatially referenced data is normally done by integrating the spatial dependence into the covariance structure either explicitly or implicitly through autoregressive models. For lattice or regional summary data, the commonly used autoregressive model is the conditional autoregressive model (CAR) or "Besag" model.

The conditional autoregressive (CAR) and the intrinsic autoregressive models (ICAR) are extensively used as prior distributions for the structured random spatial effects in Bayesian models. Some authors have pointed out unrealistic or counterintuitive consequences on the prior covariance matrix or the posterior covariance matrix of the spatial random effects (Renato *et al.*, 2009). It has been accepted also that the Besag or ICAR model may lead to ambiguous results in

the case where there is no spatial correlation in the data (Leroux *et al.*, 2000). One issue characteristic to all the models used for spatial dependency in the Besag, York and Mollie model, is that the spatially structured component is not usually scaled. It has been discussed and pointed out that scaling or weighting is vital to enable hyperprior assignment and to guarantee that hyperpriors used in one application have the same explanation in another application.

There is therefore the need to consider a weighted prior model for the spatial structured random effect. This thesis therefore seeks to propose and utilize a latent scaled Gaussian model as a prior for this spatial dependency model for flexibility, interpretability and better smoothing.

1.5 Aim and Objectives of the Study

The main aim is to carry out Bayesian spatial analyses for count data, by developing flexible and robust models for disease mapping.

Specific objectives are to:

i. to review the BYM model for disease mapping for count data and for a single disease at ecological levels.
ii. to investigate the effects of some socio-economic and demographic factors on the relative risk of tuberculosis incidence in the Province.
iii. to develop and compare some Poisson mixture models for robustness, to capture both under-dispersion and over-dispersion in a spatially correlated count data.
iv. to also develop and utilise a new and a more interpretable weighted "latent" ICAR as an alternative prior for the spatially structured random effect in the BYM model.

v. to produce a more “smoothed” graphical disease map for TB prevalence patterns by a 3-D curve fitting biharmonic spline techniques, that can suppress noise easily by seeking a least-squares fit rather than exact interpolation.

1.6 History of Tuberculosis in South Africa

Tuberculosis or TB (short for tubercle bacillus) is a popular, and in most cases, a deadly and terminal disease caused mainly by a bacteria called *Mycobacterium tuberculosis* (Kumar *et al.*, 2007). It is established that one third of the world's population is thought to have been infected with *M. tuberculosis*, with new infections arising in about 1% of the population each year (WHO 2002). More people in the developing world contract tuberculosis as a result of a poor immune system, mostly due to high rates of HIV infection and the corresponding development of AIDS (Lawn *et al.*, 2011).

Tuberculosis is termed a social disease which is caused by an airborne pathogen and has low infectivity. The spread of tuberculosis is mostly dependent on human interface within communities. However, some communities provide a better environment for the transmission for the disease than others. Earlier investigation studies have documented great differences in rates of tuberculosis among neighborhoods (Barr *et al.*, 2001). These differences may in part depend on community level, ecological risk factors that aid transmission—poverty, crowding, and other indicators of deprivation have long been associated with increased rates of tuberculosis (Hetherington *et al.*, 1929).

Tuberculosis is also closely related to both overcrowding and malnutrition, making it one of the major diseases associated with poverty (Lawn *et al.*, 2011).

The primary mode of transmission of this disease is through airborne droplets emitted (usually when coughing) by an individual who has active TB. When exposed individuals continually breathe in air contaminated by an infectious patient, they may also become infected.

Generally, the persons who are more prone to acquire infection are those in close contacts with the infected person, i.e. those who maintain long and repeated exposure to the infectious agent. About 30% of close contacts are infected, but only 15% of non-close contacts approximately (CDC and ATS, 1994). However, TB can also be spread during brief contact between individuals who do not live or work together (Small *et al.*, 2010). From history, TB has been connected with movements in population, e.g. colonization of earlier non-exposed populations, and with changes in the spatial distribution of human populations such as urbanization, i.e. increased community size and high population densities (Wilson, 1995).

South Africa is the third highest burden of tuberculosis in the world as a country, after India and China , with an estimated incidence of 450,000 cases of active TB in 2013, an increase of 400% over the last 15 years (WHO 2014). An estimated 60-73% of the 450,000 incident cases have both HIV and TB infection. The incidence of multidrug-resistant (MDR) and extensively drug-resistant TB are increasing and South Africa has the second highest number of reported multi-drug-resistant TB (MDR-TB) cases globally (NDOH 2014 and HST 2014). Tuberculosis remains the leading cause of death in South Africa, contributing to 12% of deaths in 2009 (Stats SA, 2007) as shown in **Table 1.2**. It has also been estimated that about 80% of the population of South Africa is infected with *Mycobacteriu*m bacteria, the huge bulk of whom have latent TB rather than active TB disease.

The highest predominance of latent TB, which is estimated at 88% , has been discovered among people in the age group of 30-39 years old living in townships and informal settlements (**Figure 1.4**).

Table 1.1: Statistics for TB cases notified and incidence rate of TB.

	2002	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
TB cases notified	224,420	255,420	279,260	302,460	341,160	353,870	388,880	406,080	410,040
Tb incidenc rate	493.7	550.1	599.4	645.1	719.9	739.5	798.7	823.4	802.2

Source: Notified TB cases and incidence rate of all forms of TB in South Africa (tbfacts.org 2012).

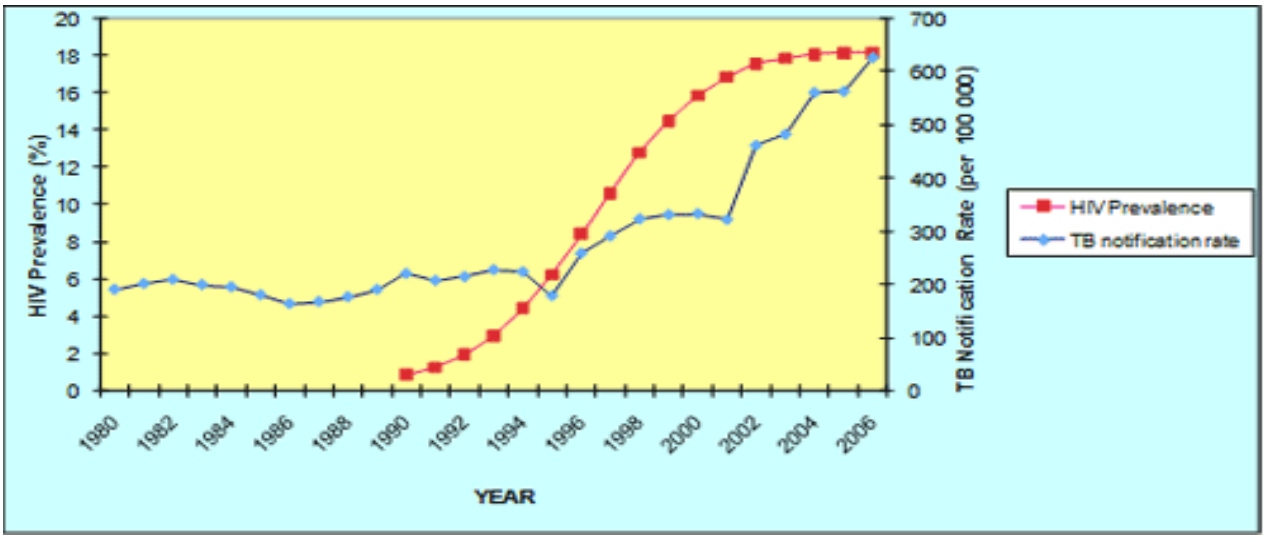
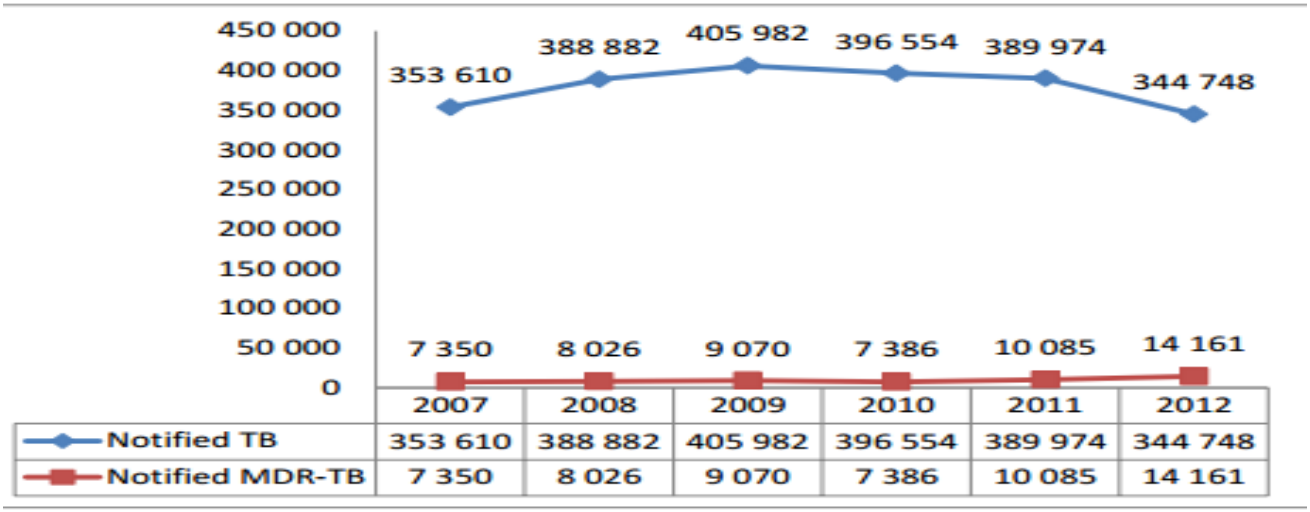


Figure 1.1: HIV and TB rates from 1980-2006.

Source:Statistics South Africa (www.statssa.gov.za).



	2007	2008	2009	2010	2011	2012
Notified TB	353 610	388 882	405 982	396 554	389 974	344 748
Notified MDR-TB	7 350	8 026	9 070	7 386	10 085	14 161

Figure 1.2: Trends for notified TB and MDR cases, South Africa (2007–2012).

Source: RSA (2014).

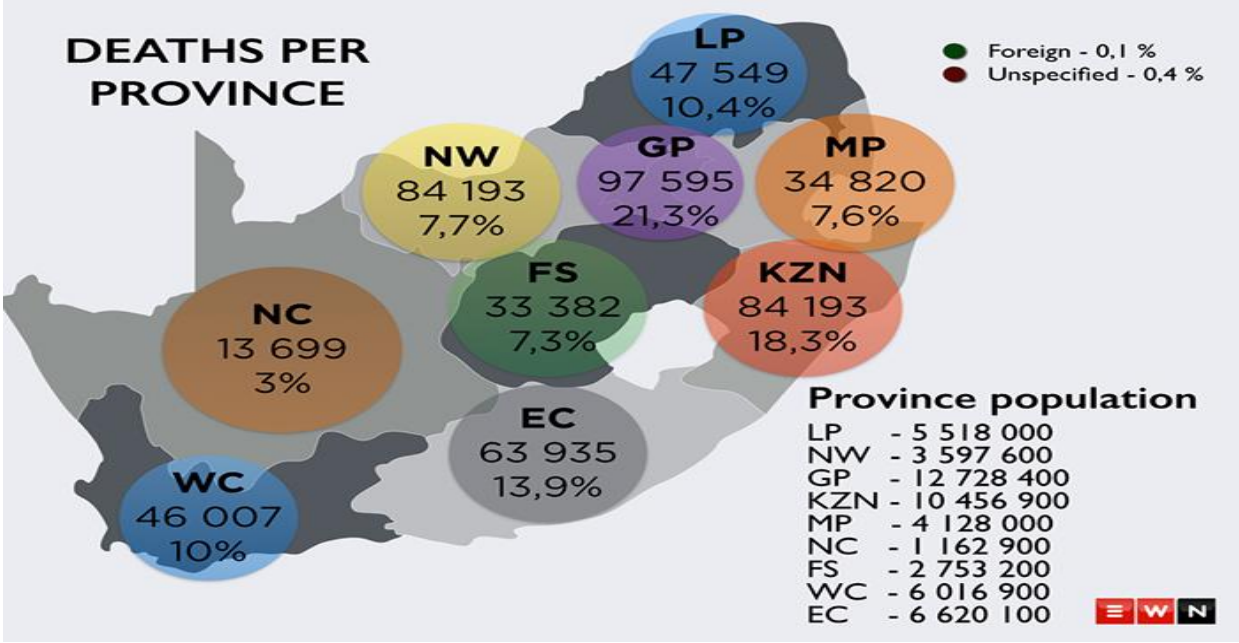


Figure 1.3: TB deaths per province.

Source: Statistics South Africa (www.statssa.gov.za).

Table 1.2: Leading cause of deaths in South Africa.

Causes of death (Based on ICD-10)	Male			Female		
	Rank	Number	%	Rank	Number	%
Tuberculosis (A15-A19)	1	4 931	11,8	1	5 975	16,8
Human immunodeficiency virus [HIV] disease (B20-B24)	2	3 230	7,7	2	4 635	13,0
Other viral diseases (B25-B34)	3	1 620	3,9	3	2 764	7,8
Influenza and pneumonia (J09-J18)	4	1 422	3,4	4	2 164	6,1
Certain disorders involving the immune mechanism (D80-D89)	5	865	2,1	5	1 478	4,1
Intestinal infectious diseases (A00-A09)	6	758	1,8	6	1 262	3,5
Inflammatory diseases of the central nervous system (G00-G09)	7	645	1,5	8	726	2,0
Other forms of heart disease (I30-I52)	8	643	1,5	7	796	2,2
Episodic and paroxysmal disorders (G40-G47)	9	473	1,1	...	...	...
Other acute lower respiratory infections (J20-J22)	10	357	0,9	10	438	1,2
Protozoal diseases (B50-B64)	...	...	...	9	604	1,7
Other natural causes		8 799	21,1		11 103	31,1
Non-natural causes		18 048	43,2		3 708	10,4
All deaths		**41 791**	**100,0**		**35 653**	**100,0**

*Excluding deaths with unknown and unspecified sex.

Source: Statistics South Africa (www.statssa.gov.za).

Table 1.3: Number and percentage of TB deaths, South Africa, 2008–2013.

Year	Number of TB deaths	% of all deaths	Tb-specific rate per 100 000
2008	**75 281**	**12.6**	**153**
2009	**69 791**	**12.0**	**140**
2010	**63 281**	**11.6**	**125**
2011	**54 112**	**10.7**	**107**
2012	**48 409**	**8.4**	**92**
2013	**40 452**	**-**	**76**

Source: SANAC (2014) and Stats SA (2014).

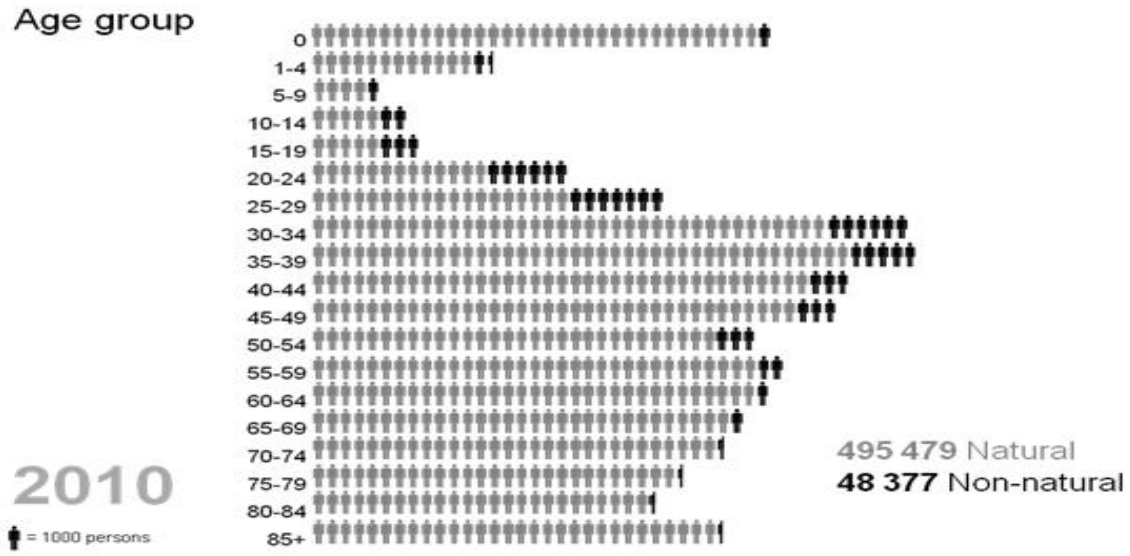


Figure 1.4: TB deaths by age group.

Source: Statistics South Africa (www.statssa.gov.za).

The scourge of tuberculosis in the Eastern Cape Province affects largely the economically active age group. For this age group of 25-34 years, the percentage distribution of reported TB cases was found to be 15.9%, 0.7% and 23.1% for the years 2003, 2004 and 2005 respectively. A report from South African National Burden of Disease study 2000 Eastern Cape Province, undertaken by South Africa Medical Research Council (SAMRC) rated tuberculosis as the second leading cause of death among women and the third leading cause of death among men aged 15-44 years.

Eastern Cape Province ranks as the second highest burden of TB by province after Kwa Zulu Natal (NDoH 2010 data) as shown in **Figure 1.3**. Also, the Eastern Cape Province has an enormously high burden of TB/HIV co-infection and MDR-TB (**Figure 1.1**). Also, there were more than 60 000 new TB cases in the province in 2008.

From these, there were 1 251 confirmed cases of MDR-TB and 385 confirmed cases of XDR-TB. In 2010, the total of new TB and re-treatment cases identified in the province stood at 62 226 (ECAC 2012).

1.7 Motivation

This research work is therefore motivated by the need to develop flexible and robust models, which extends previous studies in Bayesian spatial modelling using the Besag, York and Mollie model for count data.

Up to date, this thesis is the only analysis of the relationship between tuberculosis and transmission indicators of socio-economic and demographic variables in South Africa, despite the high burden effect of the disease. It is also one of the very few using a Bayesian disease mapping model to estimate the spatial effects and patterns of TB in the Eastern Cape province and in the country in general.

1.8 Significance of the Study

A specific challenge in epidemic modelling is the proper way to allow for spatial population structure, whereby the rate of contact between hosts depends on their spatial separation. Research of any statistical modelling of tuberculosis in the socio-demographics of South Africa has not been widely carried out. However, this would be among the first in the Province of the Eastern Cape and in the country in general. By this research, it is important to have some understanding of how a disease as serious as tuberculosis (TB) would congregate and spread through space among different regions, based on some demographic and socioeconomic risk factors in the Eastern Cape Province.

Therefore, the detection of areas with different risks for TB will take into cognizance the public health systems, and deal with the features of each region or cluster explicitly and highlight those that present higher incidences of the disease. This would also allow planning for the future, allocating scarce resources with priority and monitoring the impact of policy, political and economic changes in society and also showing the supposed direction that health policies and agencies should follow in enabling equal accessibility by all regardless of socio-economic status. The spatial effects of the spread of the disease would be determined and this would also give insight into the effects of some demographics and TB spread. Models can help propose for future zero-prevalence studies and intervention strategies by indicating where the disease frontlines are likely to be presented.

1.9 Outline of the Study

This entire thesis is about the modification of the Besag, York and Mollie (Besag *et al.,* 1995) model for disease mapping and the development of multilevel models for the spatial analysis of disease incidence using the Bayesian approach. Furthermore, this dissertation shall adopt the following outline in the course of the research.

Chapter one introduces the general concept of spatial modelling in epidemiology and some of the challenges attributable to the BYM disease mapping model for a single disease, with the common probability density (Poisson) mainly employed in spatial modelling of disease counts. The chapter also describes the rationale behind the study and the following objectives that would be investigated.

In chapter two, the basic concept of Bayesian spatial modelling in its most common manifestation was addressed. It also seeks to examine both the theoretical and empirical review of previous works surrounding the subject of the research.

Chapter three will address the method of data collection and the theories and concepts of the new methodologies that will be used to implement our objectives and in addressing our research problems. This is the contribution to knowledge that this entire dissertation seeks to perform.

Chapter four is an applied study using the Besag, York and Mollie model, addressing the spatial analyses of tuberculosis for our target regions in Eastern Cape Province based on our objectives and the overall aim of the research work. It also addresses the incidence and distribution of the disease, using the various methods sequentially as outlined in the objectives.

Chapter five will draw a general and comprehensive conclusions on all the analyses performed in the preceding chapters. Perceived shortcomings and limitations will be outlined and consequentially, reasons and areas for future research will also be stated.

In conclusion, through the interplay of developing theories, analyzing empirical data, and addressing applied problems, the work contained in this dissertation establishes some measures of progress in our understanding of disease spread in heterogeneous populations with spatial random effects.

Chapter Two

Literature Review

This chapter highlights the statistical theories and methodologies used and developed throughout this thesis, and it also gives an overview of the existing literature within this area of statistics. The objective is to outline some of the techniques employed in modelling disease incidences in the Bayesian perspectives across spatial analytic methods. It also highlighted some few cases of empirical application of Bayesian spatial methods to a disease. By choosing this review, this chapter concentrated on common theories in Bayesian spatial inference, and then concede specifics of some remarkable and recent threads of development to the appropriate literature.

2.1 Introduction

Statistical data have continually been obtained at specific places either in a forest, at a specific path address, such as in a laboratory, or an exact point on the appearance array of a gene. The location in several cases can be responsible for further understanding into situations linked through the data detail, in a word, the place where we gather a quantity may notify on "how much "or "amount" we measure. Spatial statistics as a study area contains statistical techniques engaging the inference of distance and location. Various techniques of such include additions of regression, generalized linear models, and time series as accustomed methods, whereas some arise from specific stochastic methods in space (Lance, 2005).

Substantial initial efforts in spatial analytical methods appeared in Moran (1948, 1950) and Whittle (1954). These were subsequently introductory work of spatial prediction in Matheron (1963) and Gandin (1963), also models for spatial autoregression in Besag (1974), and spatial point processes (Ripley, 1977; Diggle, 1983).

Another increasing importance in the improvement of spatial analysis is the contrasts in Ripley (1981) and Cressie (1993) writings that seek to address spatial statistics in its entirety, additionally to the current and quick growth in versions specifying certain ways of use and/or theory (Stein *et al.,* 1999; Chil`es and Delfiner, 1999; Lawson and Denison, 2002; Webster and Oliver, 2001; Waller and Gotway 2004).

The current upsurge of application in this field of spatial statistics can mostly be connected comparatively to the concurrent growths for both spatial data handiness, together with precise locality dimensions through global positioning system (GPS) technology, and speedy developments in computational power permitting improvements in the structures of spatial data (mainly within the geographic information systems) and for the insertion of extremely difficult and multifaceted models through refined algorithms and procedures. Related advancements in the speed of computations and algorithm development has moved the increase in using Bayesian approaches in this state, particularly for acknowledging and additional progress of Markov chain Monte Carlo (MCMC) algorithms, permitting huge elasticity for fitting of spatially structured statistical models.

The field of study of spatial statistics includes statistical techniques engaging place and distance in its inference. Spatial dynamics can be defined as the categorization of changes in space and time. The study of spatial analysis can also be a set of techniques which fluctuate with the locations of the objects under observation (Longley *et al.*, 2005). The spatial and the temporal process are one and the same and they cannot be separated (Shanthi and Rajan, 2012). Spatial data analysis is defined also as a collection of techniques suitable for evaluating ‘events’ at a selection of spatial scales, and the results of which can be influenced by spatial arrangement of the ‘events’ (Haining, 1994).

Significant initial study in spatial analysis showed in the studies of Moran (1948, 1950), Whittle (1954). The aim of spatial modelling is to develop a significant illustration of actions, incidences or procedures by exploring the power of spatial analysis. The two vital terms for GIS science are spatial analysis and spatial modelling.

2.2 Spatial Epidemiology Applications in Public Health Research

The study of epidemiology deals with the interrelatedness between the person, place and time in the distribution of diseases in the population (Last, 2001). In the health field, interest in spatial epidemiology began with the appreciation of maps as useful tools for understanding likely causes of disease and areas of high risk. Throughout the ages, there are instances where diseases have been wiped out after their spatial patterns were discovered. One classical example of spatial epidemiology is the London's cholera epidemic by Dr. John Snow in 1854 (Snow 1855). Snow's belief that cholera was transmitted through drinking water was refuted by some. Brody (Brody *et al.,* 2000) gave a detailed history of how Snow and others demonstrated the role of maps in the investigation of the cholera outbreak of 1854. Other examples of spatial epidemiology include the study of rickets by Palm in 1890 (Palm, 1890), and Blum in 1948 (Blum, 1948) who disclosed that sunlight was a causal factor for skin cancer. More recent studies have used spatial Bayesian techniques in malaria mortality, all-cause and cause specific mortality (Sartorious *et al.,* 2010; Musenge *et al.,* 2011). Public health data analysis has played a significant role in the development of spatial statistics in the last two decades. In spatial epidemiology, disease clusters are found for planning health care delivery and for identifying causes of the disease (Shannon 2008).

Epidemiology is also the study of how commonly diseases occur in different groups of people and why. It can also be defined as the study of the occurrence of diseases in relation to their explanatory risk factors. An important feature of epidemiology is the measurement of disease outcomes in relation to a population at risk. The population at risk is related to the group of people who are healthy or sick and counted as cases if they have the disease being studied. Epidemiological information is often used to plan and evaluate strategies to prevent an illness, and it also helps as a guide to the management of patients, in whom this particular disease has already developed. Spatial epidemiology is defined as the description and analysis of geographically indexed health data with respect to demographic, environmental, behavioural, socio-economic, genetic and infectious risk factors. In considering an analytic framework for spatial epidemiologic analyses (Elliott *et al.,* 2000b), it is vital to first differentiate between point and areal data. Each of the population, environmental exposure, and health data may be connected with a point, or a specific spatial location, for example, a street address (occurrence data), or an area, a defined spatial region such as a community, of which it is descriptive (aggregate summaries, e.g. count data). Data from a variety of these points (e.g., residence, workplace, schools) may give the closest link to an assumed biologic model in which the average disease risk of an individual will reflect individual's characteristics such as age, sex, and genetic factors (e.g., predisposition, susceptibility, immune or toxicological response capability); lifestyle variables, such as smoking and diet; and exposure to environmental pollutants (Paul *et al.,* 2004).

There are three types of spatial data that exist: areal (lattice) such as village level, point patterns (locations not fixed but random) and point-level (geo-statistical, locations are fixed over a continuous space) such as household location data (Gemperli, 2003).

Areal models use neighbourhood structures based on the arrangements of the blocks in the map. Two very popular models that incorporate such neighbourhood information are the simultaneously and conditionally autoregressive models (abbreviated SAR and CAR respectively, originally developed by Whittle (Whittle 1954) and Besag (Besag 1974), respectively. An essential property for geo-statistical data is isotropy when the spatial correlation only depends on the distance between the two locations and not directional as with anisotropy (Zimmerman, 1993).

Spatial data analysis has involved new advances in many fields, to some extent referring to an important core of knowledge, but displaying many special features in the methods involved. Many Bayesian applications have arose in spatial epidemiology, with Lawson *et al.*, (1999), Elliot *et al.*, (2000) and Lawson (2001) providing highly improved discussions. A major element in spatial outcomes analysis is the valuation of patterns of relative disease risks in terms of possible clustering around environmental point sources, but also in terms of ecological regression of disease patterns in terms of known risk factors (Lawson, 2001).

A long history of spatial modelling has occurred in spatial epidemiology. Spatial analysis has recently been added as a powerful public health tool, for its ability of visualizing disease distribution, even with sparse data (Best *et al.*, 2005), and mapping risk factors at population level (Bailey, 2001; Bailey *et al.*, 2005). The major emphasis lies in describing behavioural relationships by regression models, whether the data are defined over regions and areas, or at the level of individual actors involved in spatially defined behaviours.

Special tools in spatial statistics have advanced our understanding of the geographic distribution of diseases and improved the focus of public health actions (Anselin, 1995; Bailey, 2001).

Some recent publications have emphasized the role of Geographical Information Systems technology in Public Health research, but its use being limited by the quality of address information, particularly those available in routine information systems. Some socio-economic factors and other vulnerabilities have also been linked as risk factors to the upsurge and prevalence of disease burdens in some geographical locations.

The challenges related to transmission of infectious disease dynamics therefore, demand the attention of all researchers. Essential research on epidemiological theory must work to improve our understanding of commonly-used transmission models, and extend them to include known heterogeneities that traditionally are ignored (Dietz and Hadeler, 1988; Diekmann *et al.,* 1990; Ball *et al.*, 1997; Keeling, 1999). Empirical research must still seek to identify important patterns of heterogeneity and describe their effects on disease spread (Longini *et al.*, 1982; Grenfell *et al.*, 2001; Bjornstad *et al.,* 2002; Gani and Leach, 2004).

Some studies of social determinants of health outcomes often fail to describe how social factors influence the physiological mechanisms that may lead to various health outcomes. On one hand, infectious diseases are an individual level relationship between the host and the pathogen (Anderson, 1998). On the other hand, there is evidence that infectious disease rates are influenced more by population patterns of exposure than by the exposure status of individuals (Koopman & Lynch, 1999).

Particularly, territory-based surveillance systems have shown that the distribution of endemic diseases is also determined by social processes that are inherently related to the space where they occur (Vieira *et al.,* 2008). Such investigations may include locally related health risk factor data such as exposures to local sources of environmental pollution and the distribution of locally varying socioeconomic and behavioral factors.

It is therefore required to address how a disease like TB infection and development to the disease, which primarily occurs at the individual level, to be related to population level (i.e. social and demographic) trends.

The dynamics of infectious diseases depend on the spatial distribution of pathogens and the host, and the probability of an encounter between them. The transmission of infectious pathogens from infected to susceptible hosts decline with increasing distance between individuals. However, in large cities which tend to be overcrowded with many highly mobile individuals, the spatial correlation generated by the transmission of infection may be disrupted depending on the degree of mixing of the population (Grassberger, 1983). Since it is known that control efforts are best designed when areas of high prevalence are known, it is also important to know areas where rates are abnormally high given the underlying risk factors.

Bayesian inference has contributed greatly to this field of research which, together with computation power, friendlier and accessible software, can be used for public health purposes (Lawson, 2001). Daiane *et al.*, (2012) employed the use of a Bayesian regression model assuming a Poisson distribution to evaluate the urban spatial and temporal distribution of TB and their relationship with social vulnerability in Sao Paulo, southeast Brazil, and the model confirmed the spatial heterogeneity of TB distribution in that region, identifying areas with elevated risk and effects of social vulnerability on the disease. The relationship between tuberculosis (TB) and socioeconomic status is well known (Souza *et al.*, 2000; Waaler, 2002). Souza (2007) by a Bayesian approach, modeled the effect of socio-economic deprivation and some transmission-related indicators to TB incidence at small area level, and their model confirmed a clustered pattern of the disease and poverty (Souza *et al.*, 2007).

2.3 Bayes and Bayesian Application in Spatial Modelling

Bayesian approach remains a method of statistical inference which has been in existence for a very long time, but its uses have remained inadequate not until current improvements in simulation and computation approaches (Congdon, 2001). In the event of any statistical modelling approach, we have a vector of observed data $y_i = (y_1, \dots, y_n)$,which are believed to have come from a probability model; $f(Y|\theta)$ with a set of unknown parameters $\theta_i = (\theta_1, \dots, \theta_n)$. The purpose of statistical modelling is to use the data to deduce the best possible estimate of the values of these unknown parameters. Considering the likelihood approach, the parameters are estimated as the value, $\hat{\theta}$, which maximizes the likelihood function, represented by $L(\theta|Y) = \prod_{i=1}^{n} f(Yi|\theta)$ where $Y_1, \dots, Y_n$ are assumed to be independent. Under this context, it is assumed that the unidentified true values of the model parameters θ are fixed, with inference based on a point estimate $\hat{\theta}$ (e.g. the maximum likelihood estimator) and the uncertainty of that estimate specified by a c% confidence interval. The description of these intervals is that, if the data were recurrently sampled and an interval constructed each time, then the c% of these intervals would contain the "true" value of the parameter.

The technique of Bayesian spatial modelling is the use of Bayesian procedure to spatial models, such as models of spatial autoregression and conditional autoregressive models. A spatial dimension of this method is the use of Bayesian methods of spatial models. The idea underlying this new dimension of analyzing spatial data is Bayes' theorem that studies and accepts together the distributions of the data and the coefficient estimates that are not known (Lesage and Pace, 2009). Bayes theorem is that mathematical design of the natural idea that our estimates should change in light of observed and experimental evidence.

Precisely, the modest Bayes theorem states the association between the conditional probabilities of two events A and B as follows (Carlin and Louis, 2000):

$$P(\mathrm{E}\,|\,\mathrm{F}) = \frac{P(\mathrm{F}\,|\,\mathrm{E})P(\mathrm{E})}{P(\mathrm{F})} \tag{2.1}$$

Where $P(E\,|\,F)$ shows the posterior probability of the event E given that event F is detected. $P(E)$ and $P(F)$ characterize the prior or marginal probabilities of event E and F arising correspondingly. Equation (2.1) is applicable to a spatial modelling framework and structure. This is possible by substituting F with A to display the observed spatial data and spatial weights matrix, and E with θ to denote the parameters to be assessed in a spatial manner. We rewrite equation (2.1) as:

$$P(\theta\,|\,\mathrm{A}) = \frac{P(\mathrm{A}\,|\,\theta)P(\theta)}{P(\mathrm{A})} \tag{2.2}$$

However, the following points should be captured: (i) an assumption in a Bayesian approach is that each parameter has a prior distribution in $P(\theta)$, which captures some reservations and tells current information before observing the data. (ii) the likelihood of attaining data A in this spatial model that has the parameters, θ, is $P(\mathrm{A}\,|\,\theta)$. (iii) Basically, $P(\mathrm{A})$ is regularly set as an unknown constant which does not include any parameters in θ (Lesage and Pace 2009). (iv) the posterior distribution of θ, after allowing for both empirical data and uncertainty is $P(\theta\,|\,\mathrm{A})$.

The method of spatial modelling is primarily connected with three concerns: estimation and inference of the parameter, specification and comparison of the model, and predict from the model. It is established that spatial modelling using Bayesian techniques can solve these issues and they still have more smart and attractive characteristic indifference to the conservative method (i.e. frequentist), toward spatial modelling (Banerjee *et al.*, 2003; LeSage and Pace, 2009; Wasserman, 2003). For instance, the spatial modelling Bayes method gives an extra fixed basis as the current information and/or uncertainties of unidentified parameters are considered. The statistical inference of the posterior distribution of the Bayesian spatial modelling is more in-built and rightly agrees to probability theory.

The core dissimilarity that exists between the Bayesian approach and the frequentist is in the manner in which the unknown parameters are treated. The frequentists method mainly adopts that the data observed come from an exact chance and likelihood model, and where the unknown parameters are fixed and incomprehensible (Carlin and Louis, 2000; Congdon, 2001). On the contrary, the method of Bayesian adopts that the unidentified parameters follow prior distributions, which engages the prior distributions to get the posterior distributions of the unknown parameters.

Techniques used in Bayesian spatial modelling combines almost all the spatial models, like the spatial lag models, the spatial error models, and the well-known geographically weighted regression, and provided that Bayesian approach can be used to estimate and analyse the statistical model. Also, this spatial modelling in Bayes does not need a Gaussian spatial process, and it is more robust and elastic in generalized linear modelling (Banerjee *et al.*, 2003).

The objectives of the model are modelling the spatial variation of a random Variable $Y(s)$ at locations $s \in D$, description of trends and spatial correlation and prediction for unobserved times locations.

Amongst the advantages of adopting and applying the Bayesian method, and of current sampling procedures of Bayesian estimation (Gelfand and Smith, 1990), are the more general and usual explanation of parameter intervals, whether referred to as credible or confidence intervals, and the ease with which the true density may be obtained. The ease and elasticity of Bayesian sampling estimation spreads to resulting or structural parameters that combines model parameters and data probably, together with practical significance in usage areas (Jackman, 2000), which in classical techniques might need the delta method.

New methods of estimation assist also in the use of Bayesian random effects models for pooling strength across set of related units. These have played a key part in applications like in the analyses of spatial disease forms, survey results from lesser domain estimation (Gosh and Rao, 1994) and meta-analysis through numerous studies (Smith *et al.,* 1995). Unlike from classical methods, the Bayesian methods allow comparison of model across non-nested options, and again the latest sampling estimation improvements have enabled new techniques of model choice (Chib, 1995; Gelfand and Gosh, 1998).

2.4 Prior Distributions

The concept of a prior distribution forms a crucial part of Bayesian inference. A prior distribution, $f(\theta)$, is that which represents all of the information which is known about the parameters, θ, before observing the data, Y.

This prior distribution could be based on information from previous studies on similar data sets or an estimate from an expert in the field, or it could simply be used to represent a position of prior ignorance.

These prior distributions can take many forms subject to the type of model and data being adopted; it is possible to choose a univariate prior for each individual parameter (assuming independence between the parameters), that is $f(\theta|Y) = \prod_{j=1}^{p} f(\theta_j)$, or a single multivariate prior for all parameters together. The parameters of these prior distributions are referred to as hyperparameters. The choice of a prior distribution will influence the posterior distribution obtained, so it is essential to make a sensible choice of prior in order to produce a good and sensible estimate for the parameters. In some cases, we may have little or no insight about the value of the parameter in advance of observing the data because choice is not always direct. In such cases, it is possible to characterize our lack of prior knowledge by assigning a non-informative or weakly prior which will have a negligible result on the posterior, thus allowing the posterior distribution, $f(\theta|Y)$, to be driven by the data rather than the choice of prior.

Prior distributions can be termed as weakly informative, if it is a proper prior but has a system that the information it gives is purposely not strong, even for whatever real prior knowledge that is obtainable. Numerous non-informative prior distributions have been proposed for scale parameters for multilevel models, comprising uniform and inverse-gamma families, in the perspective of an extended conditionally-conjugate family. Several non-informative distributions for prior for σ_a, were recommended in some Bayesian literature and software, together with an improper uniform density on σ_a (Gelman *et al.*, 2003), proper distributions like $P(\sigma_\alpha^2)$ ~inverse-gamma (0.001, 0.001) (Spiegelhalter *et al.*, 1994, 2003), and distributions which rest on the data-

level variance (Box and Tiao, 1973). Completely non-informative priors can take the form of distributions without a finite density, for example Uniform$(-\infty, \infty)$.

These are called improper priors and care should be taken when using them because in many cases, they can lead to an improper posterior distribution which makes inference impossible.

Another form of prior is called the Jeffreys prior (Jeffreys, 1946), which is considered to be invariant under reparameterisation.

These Jeffreys priors are of the form

$$f(\theta) \propto \sqrt{\det I(\theta)} \tag{2.3}$$

where

$I(\theta)$ is called the Fisher information, defined as

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f(Y;\theta)\right)^2 | \theta\right] = \int\left(\frac{\partial}{\partial\theta}\log f(Y;\theta)\right)^2 f(Y;\theta)\partial y \tag{2.4}$$

2.4.1 Classes of Spatial Prior Distributions in Disease Mapping Models.

Best *et al.,* (2005) made comparisons of some Bayesian spatial models used for disease mappings by enumerating some classes of spatial prior distributions. Earlier assessments of approaches used for spatial analysis in disease risks comprise Marshall (Marshall, 1991), who contemplates empirical Bayes together with some initial entirely Bayesian approaches for disease mapping; bit hell (Bit hell, 2000) then debates that model based and nonparametric methods for both point and areal data; a little exploratory section on mapping of disease by

Lawson *et al.*, (1999) and the current section by Richardson (Richardson, 2003), which analysed the spatial modelling application techniques in epidemiology with a highlight on connecting point and area level designs.

The study of Marshall also entails cluster recognition techniques, which are not considered in this study because their objective is unlike that of creating a class of area specific risk estimates.

The resulting generic three level multilevel models have been debated by some authors (Wakefield *et al.*, 2000; Pascutto *et al.*, 2000), and likened to a basic and natural model for disease mapping, based on collection of the fundamental individual level risks:

$$\begin{aligned} Y_i &\sim Poisson\,(C_i, E_i), i = 1, \ldots, n \\ C_i &\sim P(./\theta) \\ \theta &\sim \pi() \end{aligned} \qquad (2.5)$$

with Y_i and E_i as respectively, the observed and expected number of disease cases in area *i*, C_i is the log relative risk in area i, with $P(\cdot/\theta)$ as a suitable second stage prior distribution for the $\{C_i\}$ and θ in this second stage models are termed hyperparameter with hyperprior distributions $\pi()$. The expected counts are estimated as $E_i = \sum_j N_{ij} r_j$ with r_j as the disease rate in the reference population for strata *j*, with N_{ij} as the population at risk in area i, strata *j*. This type of model is suitable if the disease condition is uncommon, and if the following assumptions are applicable:

1) if the level of risks of individuals contrast randomly as contained by areas (i.e., they are not spatially crowded),

2) in an area if the risk related to living in area i acts proportionately on the reference point risks for each stratum, then the strata specific area risks, r_{ij}, reduce to $C_i \, x \, r_j$. A binomial model for the first stage distribution is more suitable for common disease (Knorr-Held *et al.,* 1998). The major groups of some suggested spatial prior distributions for the area specific risks $\{C_i\}$ or $\{R_i\}$ are:

2.4 .1.1 Correlated Normal priors: Jointly Specified Models

Among the most flexible distributions for indicating correlated random variables is the multivariate normal (Gaussian) distribution. Given that $C = \{C_1, ..., C_n\}$ signify the vector of area specific spatial random effects given in Equation (2.6), it can be stated that the dependence structure in terms of an n x n covariance matrix $\sum$, giving rise to a second stage prior

$$C \sim MVN(\mu, \Sigma) \tag{2.6}$$

with $\sum = \sigma^2 \Omega$ and Ω_{ij} given as the correlation between C_i and C_j. If $\Omega_{ij} = 0$, note that this suggests that C_i and C_j are slightly independent.

Given some reasons of parsimony, it is normal to identify and state the essentials of the correlation matrix as a distribution-bound function of the distance, d_{ij}, among the centroids of each pair of areas, given that $\Omega_{ij=} f\,(d_{ij,} \phi)$. Nevertheless, caution is required to make sure that the preferred function effects a positive definite covariance matrix $\sum$, and there exists remarkably a small number of parametric arrangements for which this is certain (Ripley, 1981).

There exist some limited examples that are published in the use of such multivariate normal models to disease mapping, though Cook and Pocock (1983) and Richardson *et al.*, (1992) used a frequentist design in modelling the area specific errors in ecological regression and Wakefield *et al.,* (2000), gave an application to cancer outbreak disease map in the United Kingdom (UK). Diggle *et al.,* (1998) also discussed the use of mapping to count data in the environment, even though measured at point rather than areal locations.

A significant practical constraint of these models used in disease mapping techniques, even with reasonably sized study areas (which comprise of some hundred areas) is that,, application through Markov chain Monte Carlo (MCMC) algorithms, which is very expensive computationally due to the n x n inversion of the covariance matrix at each iteration. Wakefield and Kelsall (2002) proposed a similar methodology, which is centered on postulating a random (Gaussian) field (GRF) in respect of the fundamental log relative risk surface that is continuous at the second stage of the multilevel model. By joining the GRF over areas, and by using several approximations, they achieve the multivariate normal (Gaussian) distribution for the area specific log relative risks, with the average the same as the mean of the underlying GRF and the association between areas i and j is equivalent to the average correlation between the two points randomly selected from those regions.

The method is striking given that correlation and association between areas is invoked through a point level correlation function and it is likely to restructure the posterior distribution of the principal and underlying continuous risk surface. The data nevertheless, will comprise little facts about the spatial dependence at distances which are below the size of the smallest areas, so that structures of the posterior risk surface at a small scale should not be too inferred.

2.4.1.2 Conditionally Specified (Markov random field) Models

The second stage multilevel model most frequently used in disease mapping is the Gaussian Markov random fields (MRF), by following the ingenious documents by Clayton and colleagues (Clayton *et al.*, 1987; Clayton *et al.*, 1992) and Besag *et al.*, (1991).

Typically, the models are stated by a succession of some conditional distributions, which in broad form, can be given as

$$E\left(C_i \middle| c_{(-i)}\right) = \mu_i + \sum_j a_{ij}\left(c_j - \mu_j\right);\ a_{ij} \neq 0\ , a_{ij} = 0, i = 1, \ldots, n, \quad (2.7)$$

$$Var\left(C_i \middle| c_{(-i)}\right) = k_i > 0, i = 1, \ldots, n. \quad (2.8)$$

where $c_{(-i)}$ signifies the values of the random effects in all the areas except for the *ith* area (Besag, 1974; Cressie, 1993).

Models of such are also called the Gaussian conditional autoregressions, where the μ_j parameters signify a huge and large scale spatial trend or gradient at location i, and they are generally assumed to be constant across locations or quantified as a function of covariates. Generally, we assume $\mu_i = 0 \forall_i$ here and there are a_{ij}, coefficients imitating the local spatial reliance that are between the units i and j.

With reference to some definite restraints on a_{ij} and k_i, it can then be presented that equations (2.7) and (2.8) give a type of joint distribution as

$C \sim MVN(\mu, P^{-1})$ where $\mu = (\mu_1, \ldots, \mu_n)(=0), P_{ii} = 1/k_i$ and $P_{ij} = -a_{ij}/k_i$.

Given P as the inverse covariance matrix, then it must be symmetric giving rise to the constraint $a_{ij}k_j = a_{ji}k_i$.

A broadly used design that fulfills this type of symmetry situation is

$$a_{ij} = \frac{\phi w_{ij}}{w_{i+}}; w_{i+} = \sum_j w_{ij} \tag{2.9}$$

$$k_j = \frac{\sigma^2}{w_{i+}} \tag{2.10}$$

Knowing that the term k_i must be positive, caution is required if any of w_{ij} is negative. The term w_{ij}, is given as the $ijth$ element of a symmetric n x n weight matrix W, with diagonal components, $w_{ii} = 0$. A general option is to set $w_{ij} = 1$ if locations i and j are neighbours and $w_{ij} = 0$, otherwise. The parameter, ϕ, can be assumed as an autocorrelation parameter, which reveals the complete power of spatial dependence between regions with nonzero weights.

An attractive characteristic of the MRF prior (2.8) – (2.9) is the likelihood of making an inference or extrapolation about the total amount of spatial dependence in the disease risks by approximating, ϕ. Nevertheless, explanation of ϕ is not direct (Sun *et al.,* (2000) for an explanation of ϕ as a spatial shrinkage factor), and for values near to the maximum $1/\lambda_{max}$, are desired to reveal uniform adequate spatial dependence (Besag, 1981).

Alternatively, if $\phi = 0$, it shows independence between areas, but in the case of the prior for MRF parameterized, using Equations (2.9) and (2.10) does not shrink to that typical independent hierarchical/multilevel normal prior for the log relative risks as the variance is unequal across areas.

This made Besag *et al.* (1991) to suggest a substitute for the second stage prior model (hereafter indicated Besag, York and Mollie (BYM) model)

$$C_i = V_i + U_i, \qquad i = 1, \ldots . n \tag{2.11}$$

$$V_i \sim Normal(0, \sigma_v^2) \tag{2.12}$$

$$U_i | u_{-i} \sim Normal\left(\frac{\sum_j w_{ij} U_i}{w_{i+}}, \frac{\sigma_\mu^2}{w_{i+}}\right) \tag{2.13}$$

The $\{U_i\}$ follows a central autoregression, realized by setting up ϕ to its restrictive value of 1, and likened to be signifying a between area variation of a spatial element in a disease risk. Though the conditional univariate prior distributions (2.13) are adequately specified, the equivalent joint prior distribution for U is now termed as improper (as a vague mean and infinite variance). A posterior distribution that is very appropriate will however be achieved, conditional on the standard prerequisite for proper hyperprior distributions on variance components in multilevel designs (Sun *et al.*, 1999). The $\{V_i\}$ characterized the geographically uncorrelated elements of heterogeneity in the disease risk.

The posterior implication about the degree of spatial dependence is then centered on the quantity of the total (marginal) variation in the $\{C_i\}$, covered by each element (whereby the marginal variance of the $U_i's$ is valued empirically at each MCMC iteration).

As described before, the MRF models in (2.8) and (2.9) are equal to postulating a multivariate Gaussian model for area specific random effects joint distribution, but giving distribution parameters to the dependence structure in terms of the precision matrix, P, (signifying conditional independence assumptions), reasonably than the covariance matrix $\sum = P^{-1}$ (signifying marginal independence assumptions).

Nevertheless, there are significant computational benefits to modelling the accuracy matrix, as the independence assumptions that are conditional are freely misused by MCMC algorithms, so that the MRF models can be resourcefully employed lacking the necessity for matrix inversion. MacNab (2003) recommended a correlated multivariate normal model, which is also parameterized in terms of the precision matrix. It was assumed that $C \sim MVN(0, \sigma^2 D^{-1})$ with $D = \rho P + (1-\rho)I$ by MacNab, where P is given to be the MRF model, I, is defined as the n x n identity matrix and $\rho \in [0,1]$ can be inferred as measuring for spatial dependence for the case that if $\rho = 0$ or 1, then the model shrinks to the normal (Gaussian) independence prior (2.12) or to the intrinsic autoregression (2.13), respectively.

This specific model may well therefore, be perceived as a substitute to the BYM model, but with the benefit that it evades the possible identifiability difficulty faced by the BYM prior, where only the summation of $U_i + V_i$ is fully recognized by the data (Eberly *et al.*, 2000). MacNab also reported generally related outcomes in a contrast of the two designs when used in modelling small area health service utilization and outcome rates, though her model was not as much

profound to the hyperpriors, and gave rise to a somewhat improved fit in her specific application. An alternative adjustment and broadening of the model (BYM) has been recommended by Lawson and Clarke (2002). Their model contains a combination of normal and non-normal (median based) conditional autoregressive elements, with the non-normal aimed to identify and capture discrete jumps in the relative risk structure in a bid to escape over smoothing of the risk surface.

2.4.1.3 Semi-parametric Spatial Models

Best *et al.*, (2005) suggest that in the case of the parametric requirements or specifications defined earlier, the extent of smoothing executed in the model (BYM) is disturbed generally by all the regions and is not adaptive.

With a concern that the parametric models of such, could over smooth the relative risk surface have made numerous authors to improve on semi- parametric spatial models, to substitute the constantly changing spatial distribution for $\{C_i\}$, by discrete distribution or partition models within each cluster area or component with a constant unknown relative risk. As one of the common features, these models permit gaps in the risk surface and they make fewer assumptions about the distributions, though implementing however, a required quantity of smoothing and borrowing of strength by letting areas to be assigned to the same cluster. These models are different in the way that they allot areas to cluster groups or to the mixture components. This group of models possess the possibility to over fit the data by generating a boundless number of needless clusters, and it is likely to reason it is necessary to add penalization that is categorical of the model measurement.

Actually as debated by Denison *et al.*, (2002), the Bayesian framework with functional selection of hyperpriors for the model parameters will inevitably favour a simpler model over a more multifaceted one, if the data is sustained by the former, since the prior weight will be further concerted about the observed data for the lesser dimensional prior and hereafter, the modest and simpler model will have a greater posterior probability

2.4.1.4 Spatially Dependent Mixture Models with Allocations

A mixture model suggested by Green and Richardson (2002), for the C_i values states that the distribution of each area to a risk grouping sticks to a spatially correlated procedure. This mixture type of model can be viewed as an extension lead of hidden Markov models, to unknown discrete state MRF, namely, MRF lowered by (conditionally) independent noise. The distribution or allocation model selected is called the Potts model, commonly used in image processing, and consists of an interaction parameter that guides the extent of spatial dependence. The amount of the constituents of this mixture is not defined before, and it is also estimated as part of the model.

Therefore, the model is given as follows:

$$\exp(C_i) = \eta_{zi}, \quad i = 1, ..., n$$

z_i (allocation/distribution variable) $\in \{1, ..., k\}$ (selected according to Potts model)

$$p(z \mid \lambda, k) = e^{\lambda u(z) - \sigma k(\lambda)} \text{ (Potts model)}$$

$$\eta_j \sim Gamma(\alpha, \beta) \; i = 1, \ldots, k$$

$$k \text{ (the amount of components)} \sim Unif(1, e_{max})$$

where $\lambda > 0$ is the parameter of interaction to be estimated and $U(z) = \sum_{i-i'} I\left[z_i - z_{i'}\right]$ is given as the amount of characterized pairs of the neighbouring areas.

Hence, when a region i is allocated to a component j it will be preferred by taking neighbouring regions in the same component j, and the more it is so, the greater the value of c.

This procedure of allocation not only takes prior beliefs about spatial resemblance of threats in nearby areas, but it also permits noncontiguous areas to fit into similar element.

In the Potts model, the last component $\sigma_k(\lambda) = \log(\sum z \in \{1, 2, \ldots, k\}^n e^{\lambda u(z)}$ is called the normalizing constant, where the totality is the complete probable structures of the allocation for the n areas.

2.4.1.5 Spatial Partition Models

These authors, Knorr-Held and Raßer (2000) and Denison and Holmes (2001) introduced a set of closely linked semi-parametric models and they called them spatial partition models. For any of these partition models, it is anticipated that there would be a group of *k* non-overlapping clusters of areas, and each of them with uniform relative risk, and the *k* term is given as unknown. These models are different technically in the manner the clusters are defined and in their hyperprior condition.

For Knorr-Held and Raßer (2000) model mentioned above, the term k of the regions are chosen at random as the supposed cluster "hot spot" centres $g_j, 1 \leq j \leq k$. With these restrictions, the left over areas are assigned to cluster j if its center is closer than any other in terms of the insignificant number of region boundaries that have to be cut across to get to it. Also restricted on the term k is each structure of centres assumed to be equally likely a priori.

The following model is then given:

$$C_i = \eta_{zi}, \qquad i = 1, \ldots, n$$

$$z_i \in \{1, \ldots, k\} \text{ (ideal according to allocation procedure defined earlier)}$$

$$\log \eta_j \sim Normal(\mu, \sigma^2) \qquad j = 1, \ldots, k$$

$$k \text{ (amount of clusters)} \sim Unif(1, e_{max}) \text{or geometric.}$$

Therefore, knowing that all the regions in a cluster are joined, while in the spatial mixture model earlier defined, the regions in the same part are not essentially connecting. Therefore, the amount of *k* clusters, in the KHR's model will be likely to be much greater than the number of elements in the mixture model. In assigning areas to a cluster group, the duo Denison and Holmes adopted the hint given by Voronoi tessellations as a clustering tool.

Every single region is considered as a point location and characterized \by means of the co-ordinates of its centroid. The tessellation of Voronoi points of generations are presumed to be situated at any point in the area of concern. Restrictive on a set of k creating points (similar to the g_j in KHR's model), partition of the jth component is then comprised of regions with centroids nearer (in Euclidean distance) to the jth generating point than to any other point.

Another spatial partition model that is associated is that established by Gangnon and Clayton (2000). This model particularly, also assemble regions into clusters, but the principal emphasis is not on the flexible and supple modelling of the hazard or risk surface, but instead, the making of extrapolation about the number, location and configuration of clusters.

These are being modeled as placed over, on a varied related risk surface by making use of an exact group of priors for their structure.

2.4.1.6 Spatial Moving Average models

There exist classes of flexible models called spatial moving average models that have been applied and used to define a continuous spatial process, generally in presentations of geostatistics. Models of such are developed by incorporating a simple two dimensional random noise procedure (e.g., a grid of iid normal random variables) with a particular leveling kernel which is a function of distance and probably, location. This kernel can be supposed to be a method to depict the random noise process in the two dimensional space in order to give a smooth surface.

The benefit of this spatial moving average technique over straight modelling of the covariance function lies in the fact that rich groups of kernel functions can be taken into account for modelling exact descriptions (e.g., nonstationarity, edge effects) of the spatial dependence structure (Higdon 2002) although still conserving the features of the fundamental covariance function. These models were principally developed for continuous procedures, and so they have just little application to a disease mapping setting.

Best *et al.*, (2000; 2002) however, suggested a discrete form of a gamma moving average procedure in modelling geographical differences in childhood respirational diseases. Their model was founded on the additive risk models of (2.15) and (2.16):

$$Y_i \sim Poisson\ ((ro + R_i) \times M_i), i = 1,...,n \quad (2.14)$$

$$R_i \sim P(.|\theta),\ R_i > 0 \quad (2.15)$$

where $M_i = \sum_j \beta_j N_{ij}$ and can be assumed to be the 'standardized' population in area and the region specific random effects R_i denote unmeasured spatially fluctuating extra risks.

For each R_i the next stage of the model is built by indicating a random grid of latent *iid* gamma random variables γ_j (($j = 1,...,m$) where m is denoted the total number of grid cells describing the latent process) hiding the study area.

An isotropic, stationary Gaussian kernel function was assumed by Best and partners, though other kernel methods are basically adapted. Formally, the second stage model is given as follows:

$$R_i = \sigma \sum_{i=1}^{m} k_{ij} \lambda_j, \qquad i = 1,...,n \quad (2.16)$$

$$\lambda_j \sim Gamma(\alpha_j, \tau), \quad j = 1,...,m \quad (2.17)$$

$$k_{ij} = \frac{1}{2\pi\phi^2} e^{-d_{ij}^2/2\phi^2}$$

where σ can be defined as a scale factor for the spatial random effects, the distance between the centroid of area i is given as d_{ij}, while the latent grid cell centroid given j and ϕ is defined as the spatial range parameter overseeing how speedily the effect of the latent gamma random variables with distance on the area specific extra risk decreases.

An attractive explanation of this model is to picture the gamma random variables as representative of the location and size of unmeasured risk influences, and the region exact random effects as demonstrating the cumulative outcome of these risk factors in each region, weighted by their distance from the region fitting to the kernel 'weights' k_{ij}. Nevertheless, this model ensures more fine tuning other than the semi-parametric or that of the BYM models, since the number and extent of the latent grid cells must be quantified ahead by the user, and it is not at all times clear on how best to select these.

Best *et al.,* (2000) deliberated in full the hyperprior description. However, note that the prior shape (α_{ij}) and the precision (τ) parameters of the latent gamma variables ought to be selected in such a way that the term g_j has prior mean relative to the part of the jth latent grid cell (thus the subscript for γ_j to get used to nonregular grids). This enables the model to be spatially extensible with the knowledge that any division of the latent gamma random variables will give rise to same probability distributions for the kernel weighted sums in Equation (2.17). Since the undetected risk factors are characteristically defined on a reduced geographic partition other than the disease outcome data, maps of the posterior risk surface can be created for other geographical partition, just by estimating the kernel sum (2.17) at these needed places. However, a possible restraint is that the model depend on the theory of additive risks which may not continually hold in reality, and it does not assure that the Poisson degree in each area, $(r_0 + R_i)$, will be assessed to be fewer than 1 (though this is not likely to be a problem as long as the disease is uncommon).

2.5 Posterior Distributions

The distribution of the parameters after detecting the data is defined as the posterior distribution. This distribution (posterior) is attained when the prior distribution with the observed data is updated. A full Bayesian inference attains its estimations by sampling from this posterior distribution. The posterior distribution in certainty is generally of high measurement and dimension, and it is also systematically intractable. This is worsened by the weighty and tedious integration essential when presenting analytical approaches.

The Markov chain Monte Carlo (MCMC) approaches is a class of methods that have been used to surmount this problem. The MCMC permits for a straight sampling from this posterior distribution repetitively and approximations are calculated from these samples by means of simple data extractions such as the mean and median (Ngesa, 2014).

2.6 Multilevel/Hierarchical Models

Multilevel models are also known as hierarchical models for two different reasons: firstly, from the way the data are structured and secondly, the model itself, which possesses its own exact hierarchical structure. These multilevel/hierarchical models are regularly known as random-effects or mixed-effects models. Hierarchical models are used when the data are structured in groups for example, demographically, temporally, spatially or when different (but related) parameters are used for each group. These multilevel Bayes models are actually the mixture of two things: (i) when a model is written in a hierarchical procedure that is (ii) when such a model can be estimated using Bayesian techniques.

Multilevel regressions usually have different intercepts or slopes (or both) for each group and are constrained by a distribution – often a normal with a variance that we model (Gelfand *et al.,* 1990).

The one that is written modularly is a hierarchical model, or when it is in terms of sub-models.

It is beneficial to consider the analysis of epidemiologic data using a particular model for the within-unit analysis, and use a different model for across-unit analysis. When describing the behavior of individual/aggregate respondents over time, the within-unit model could be used, while when describing the diversity, or heterogeneity, of the units, the across-unit would be appropriate. The combination of the sub-models form the hierarchical model and the pieces are integrated by the Bayes theorem and which also account for all the uncertainty that is present (Greg *et al.,* 2005).

Hierarchical models are fundamental and central to contemporary Bayesian statistics both for the reasons of conceptuality and practicality. On the hypothetical basis, hierarchical or multilevel models permit a more objective and unbiased methodology to inference, by estimating the prior distribution parameters from the data instead of requiring that they be indicated by means of subjective evidence and information (James and Stein, 1960; Efron and Morris, 1975; and Morris, 1983). At the real-world level, multilevel models are flexible and robust tools for joining information and fractional inferential pooling (Kreft and De Leeuw, 1998; Snijders and Bosker, 1999; Carlin and Louis, 2001; Raudenbush and Bryk, 2002; Gelman *et al.,* 2003).

In the (generalized) linear models, the observations are independent of each by assumptions given the predictor variables. Although, there are some circumstances in which independence of that type does not stand. A main type of condition that infringes these assumptions of independence is the cluster-level characteristics.

This happens when observations belong to dissimilar clusters and when each of the clusters has its own features (Roger *et al.*, 2012).MCMC methods have been known to function mostly well with multilevel models, and this has powered the development and application of Bayes' theorem.

Hierarchical or multilevel models is when a tasking difficulty is broken into a sequence of levels connected by non-complex rules of probability, when it assumes a very elastic and easy structure proficient of accepting ambiguity and possible a priori scientific information while it still retains many benefits of a stringent likelihood method for example, many sources of data and information with scientifically significant structure (Ali *et al.*, 2006).

When complicated procedures are modeled in the presence of data, let θ_1, θ_2 and θ_3 be data, process and parameter models respectively, it is useful to present the multilevel or hierarchical model in three phases basically:

$$Stage\ 1.\ Data\ Model: \left[data \mid process;\ data\ parameters\right]$$

$$Stage\ 2.\ Process\ Model: \left[process \mid process\ parameters\right]$$

$$Stage\ 3.\ Parameter\ Model: \left[data\ and\ process\ parameters\right],$$

So that, when making use of the Bayesian method, the joint or posterior distribution of the process and parameters when the data is given, is hereby acquired through the Bayes' Theorem:

$$\left[process;\ parameters \mid data\right] \propto \left[data \mid process;\ parameters\right]\ x\ \left[process \mid parameters\right]\left[parameters\right]$$

The first phase is connected to the observational procedure or the model of the data, which states the distribution of the data, when given the central process of concern and parameters that define the data model. The second stage goes ahead to describe the process, which is conditional on other process parameters. Lastly, the uncertainty in the parameters is modeled by the last stage, from the combination of the data and process phases. Also know that each of these steps can have several sub-stages (Wikle *et al.,* 1998, 2001). The overall objective is to quantify by estimating the distribution of the process and parameters by the given data.

The methods of Bayesian are logically suitable for approximation in such hierarchical situations. A multilevel model needs hyperparameters, and these must be assigned their specific prior distribution (Gelman, 2006). The priors considered above for the underlying population mixing density have an exact parametric form. Various non-informative prior distributions for σ_α have been proposed in some Bayesian documents and software, together with an improper uniform density on σ_α (Gelman *et al.,* 2003), proper distributions such as $p\left(\sigma^2{}_\alpha\right) \sim$ inverse-gamma(0.001, 0.001) (Spiegelhalter *et al.,* 1994, 2003), and the distributions that rely on the data-level variance (Box and Tiao, 1973) as indicated before.

2.7 Generalized Linear Models

Generalized linear modelling is the background for most analysis in statistics that contains linear and logistic regression as distinct cases. A generalized linear model usually contains:

1. A data of vector $y_i = (y_1, \dots, y_n)$
2. Predictor variables X and coefficients β , creating a linear predictor $X\beta$
3. A specified link function g, that yields a vector of transformed data $\hat{y} = g^{-1}(X\beta)$ that are utilised to model the data

4. A type of data distribution , $p(y \mid \hat{y})$
5. Probably other parameters like variances, over-dispersions, and cutpoints contained in the predictors, link function and data distribution.

The choices made for a generalized linear model are the link function *g*, and the specified data distribution, p.

Considering the statistical modelling employed for this research, the data usually consists of an outcome variable$Y_i = (y_1, \ldots, y_n)$and a given set of covariate data $X = \left(x_1^T, \ldots, X_n^T\right)$,

where $x_i^T = \left(x_{i1}, \ldots, x_{ip}\right)$, the set of p-covariate values linking to observation *I* and x_i is a vector of ones for the intercept term. The objective of the modelling methodology is to estimate a set of regression parameters $\beta = (\beta_1, \ldots, \beta_p)$, which best define the relationship between the responses and these covariates.

The well-known and simplest modelling approach is the linear model, which represents a direct association between the covariate data and the response variable, and takes the form:

$$Y_i \sim N(\mu_i, \sigma^2), \text{ i } = 1, \ldots, n \tag{2.18}$$

$$\mu_i = x_i^T \beta \tag{2.19}$$

where each response Y_i is presumed as an independent Gaussian random variable with mean and variance μ_i and σ^2 respectively.

For the Bayesian framework, we can assign a prior $\beta \sim N(0,\sigma_b^2)$, to indicate a lack of any strong prior certainty about the intercept, and a conjugate prior $\sigma^2 \sim InvGamma(\alpha,\psi_1)$, and then the full conditionals are given as:

$$f\left(\beta \mid \sigma^2, Y\right) \propto \prod_{i=1}^{n} N\left(Y_i \mid x_i^T \beta, \sigma^2\right) \times \prod_{i=1}^{p} N\left(\beta \mid 0, \sigma_b^2\right)$$

$$\propto \prod_{i=1}^{n} \exp\left(-\frac{\left(Y_i - x_i^T \beta\right)^2}{2\sigma^2}\right) \times \prod_{i=1}^{p} \exp\left(-\frac{\beta_j^2}{2\sigma^2}\right)$$

$$f\left(\sigma^2 \mid \beta, Y\right) \propto \prod_{i=1}^{n} N\left(Y_i \mid x_i^T, \sigma^2\right) \times \prod_{j=1}^{p} invGamma\left(\sigma^2 \mid \alpha, \psi\right)$$

$$\propto \left(\sigma^2\right)^{-\frac{n}{2}} exp\ \{\ (\ f\ (\beta \mid \sigma^2,\ Y) \propto \prod_{i=1}^{n}\ N\ (Y_i \mid x_i{}^T \beta, \sigma^2)\ x \prod_{j=1}^{p}\ N\ \left(\beta \mid 0, \sigma_b^2\right)$$

$$\propto \prod_{i=1}^{n}\ exp\left(-\frac{(Y_i - x_i^T \beta)^2}{2\sigma^2}\right) \prod_{j=1}^{p}\ exp\ \left(-\frac{\beta_j^2}{2\sigma_b^2}\right)$$

$$f\ (\sigma^2 \mid \beta,\ Y) \propto \prod_{i=1}^{n}\ N\left(Y_i \mid x_i^T \beta, \sigma^2\right) \times \prod_{j=1}^{p}\ invGamma\ (\sigma^2 \mid \alpha, \psi)$$

$$\propto \left(\sigma^2\right)^{-\frac{n}{2}} exp\left(\frac{\sum_{i=1}(Y_i - x_i^T \beta)^2}{\sigma^2}\right) \times (\sigma^2)^{-(\alpha+1)} \exp\left(\frac{\psi}{\sigma^2}\right)$$

$$\hat{\theta} \propto \left(\sigma^2\right)^{-(\alpha+\frac{n}{2}+1)} \exp\left(\frac{\sum_{i=1}(Y_i - x_i^T \beta)^2 + \psi}{\sigma^2}\right)$$

$$\sim InvGamma\left(\alpha + \frac{n}{2}, \psi + \sum (Y_i - x_i^T \beta)^2\right) \qquad (2.20)$$

A generalized linear model (GLM), Nelder and Wedderburn (1972) is an extension of this linear model form, which allows for more flexibility and robustness in the modelling approach.

For this approach, there is no longer a necessity for it to be a direct linear association between the outcome variable and the covariates, and this outcome or response variable, Y , can be any set of independent random variables from any exponential family distribution, f . For some random variable, Y , and some parameters, η, this exponential family is a set of statistical distributions which, accepts that $f(y|\theta) = exp\,(a(y) + b(\theta) + c(y) + d(\theta)$, where, a, b, c, d are a set of recognized functions. Some known members of this exponential family are the Gaussian, Binomial, Exponential, probit, multinomial logit and Poisson distributions.

A generalized linear model therefore takes the form:

$$Y_i \sim f(\mu_i) \qquad i = 1, \ldots, n$$

$$\mathrm{h}(\mu_i) = \eta_i = x_i^T \beta \tag{2.21}$$

From (2.21), $\eta_i = x_i^T \beta$, is identified as the linear predictor, and h() is a recognized monotonic invertible function called a link function. Common examples of the link function h() include log, square root and logit transformations. Note that the linear model defined above is known as a distinct case of the GLM that is obtained where the link utility or function is simply the identity function

$$\mathrm{h}(\mu_i) = \mu_i \text{ and } f(Y_i|\mu_i) = N(\mu_i, \sigma^2) \tag{2.22}$$

2.8 Poisson Generalized Linear Model (GLM) For Count Data

Poisson regressions have been extensively used to model count data. It is also used for modelling the dissimilarity or variation in count data (that is, data that can be equal to $0, 1, 2, \ldots,$). Nevertheless, it is evaluated regularly for its constricting assumption of equal dispersion between the variance and the mean. But in real life and practical applications, count data always show over-dispersion (Osei, 2010).

We can calculate probabilities for counts through the theoretical formula that follow a Poisson probability function which is characterised by a single parameter, λ. Suitably, λ is the theoretical mean and estimated by the sample mean under both likelihood and moments. Consequently, if we have an estimated value for λ, we can immediately calculate the corresponding probability that a count equal to Y_i is observed in a Poisson distribution with mean, λ. Considering modelling the count of people with a particular disease within a set of areas $i = 1, 2, \ldots, n.$

The problem can be measured from a hierarchical viewpoint, where the disease count data is modeled as Poisson. We introduce random effects to account for spatial dependence unexplained by the observed data. For spatially indexed data, the random effects linked to each area or location may be explored, giving way for the modelling of a fundamental spatial dependence structure. It is preferred considering the modelling in the frame work of the Bayesian approach. This permits adjustment of the uncertainty parameters by allocating parameters with prior distributions. Structures for spatial correlation are then integrated in a Bayesian background by assuming appropriate spatial random effects prior distribution.

Models for Bayesian approximation of the Normal straight line regression model, whether with univariate or multivariate result are well recognized. If we assume that the predictors are exogenous and measured without error, that is, not random, they may be conditioned on a fixed constants. For a parameter with univariate outcome, the regression coefficients link the mean outcome for case I to predictors $\sigma_{\theta 2}$, for that case and the conditional or residual variance.

General Linear models (GLMs) have been proposed as an integrated structure for both types of outcomes. McCullough and Nelder (1989) elaborated that several discrete densities can be included within the exponential family. The exponential family density (Gelfand and Ghosh, 2000) has the form:

$$f(y_i \mid \theta_i) = exp\left\{\phi_i^{-1}\left[y_i\theta_i - b(\theta_i)\right] + c(y_i, \phi_i)\right\} \tag{2.23}$$

where $\phi_{i,}$scale parameters, and the means are obtained as:

$$E(y_i) = \mu_i = b(\theta_i) \tag{2.24}$$

and variances as

$$V(y_i) = \phi_i b(\theta_i) = \phi_i V(\mu_i) \tag{2.25}$$

Poisson and binomial densities have a fixed scale factors $\phi_i = \phi = 1$. So for the Poisson with

$$b(\theta_i) = exp(\theta_i), \tag{2.26}$$

the mean and variance are both μ_i. Thus for a Poisson outcome, $Y_i \sim Poi(\mu_i)$ a model is stipulated for the mean μ_i containing both fixed and random effects:

$$\log(\mu_i) = \beta X_i + \varepsilon_i \tag{2.27}$$

where ε_i are parametric (Normal) or possibly non-parametric. This well means that adding a set of parameters, which increase in number with the sample size, theoretically making the likelihood nonregular and raising the question about how many parameters are actually in the model (Congdon, 2003).

The Poisson distribution belongs to a known exponential family of distributions, and so it is a generalized linear model that can be used with these types of data. Therefore, the response data from the Poisson distribution can only take non-negative values, so the log is an appropriate and commonly used link function which allows that the model always fits non-negative values.

The basic Poisson GLM can take the form as follows:

$$Y_i \sim Poisson\ (\mu_i) \tag{2.28}$$

$$Log\ (\mu i) = x_i^T \beta \tag{2.29}$$

In the Bayesian framework, if we assign the prior $\beta \sim N\ (0, \sigma^2{}_b)$ then the full conditional is given as:

$$\mathrm{f}\left(\beta \mid \sigma^2,\ Y\right) \propto \prod_{i=1}^{n}\ Poisson\ (x_i^{\ T}\beta) \times \prod_{j=1}^{p}\ N\left(\beta \mid 0,\ \sigma^2\right)$$

$$\propto \prod_{i=1}^{n}\ exp\ (-x_i^{\ T}\beta)\ \left(x_i^{\ T}\beta\right)^{Yi} \times \prod_{j=1}^{p}\ exp\left(-\frac{\beta_j^2}{2\sigma^2}\right) \qquad (2.30)$$

2.9 The Besag, York and Mollie (BYM) Model

The most commonly used model in single disease spatial analysis was proposed by Besag *et al.*, (1991). The model was used for disease incidence and prevalence in areas by means of the Poisson model. Areal count data are usually modeled by extending the Poisson log-linear model to explain for the spatial form of the disease data.

The data is possible to contain spatial autocorrelation, where correlation exists between pairs of areal units which are close to each other geographically. The spatial pattern of the data is modeled by a combination of covariate data, $X = \left(x_1^T, \ldots, x_n^T\right)$ and a set of random effect terms, $\theta = (\theta_1, \ldots, \theta_n)$. These random effect terms usually represent the unexplained spatial autocorrelation and some unmeasured confounding and/or latent variables brought into the disease data.

The spatial models commonly used with count data ***Y*** are typically Poisson GLMs of the form

$$Y_i = Poisson(\mu_i) \qquad i = 1,...,n, \tag{2.31}$$

$$\log(\mu_i) = x_i^T \beta + \theta_i$$

These random effects $\theta = (\theta_1, \ldots, \theta_n)$ are usually modeled using the widely known conditional autoregressive (CAR) prior. These types of models can be indicated by a group of univariate full conditionally restrictive distributions of the form $f(\theta_i \mid \theta_{-i})$, where $\theta_{-i} = (\theta_1,...,\theta_{i-1},\theta_{i+1},...,\theta_n)$.

Besag *et al.,* (1991) in addition, modified this model further by the addition of a spatially unstructured random effect and a spatially structured random effect, through what is called the convolution model. This is to permit the model to use information both at the local and global level. Also the need arises to allocate prior weights equally to these two model components in order to evade either the global over smoothing or the local over smoothing. The BYM model was used for disease dominance and incidence in areas using the Poisson model.

For the model, let the relative risk unknown for area i with regard to a regular population be defined as λ_i. Let Y_i also indicate the counts of disease observed in region i and let e_i indicate the count expected in the same location. The log of the relative risk by assumption has it, that the disease can be split into two components namely: spatially structured component u_i and a spatially unstructured v_i.

This can then be mathematically given as

$$Y_i \sim Poisson\ (e_i, \lambda_i) \tag{2.32}$$

with

$$log\ (\lambda_i) = u_i + v_i \tag{2.33}$$

where μ_i and v_i denote the random effects signifying the unobserved covariates, with μ_i signifying variables that if they were detected would impact on the spatial structure, while v_i characterizes the unobserved heterogeneity in region *i*. Besag *et al.,* (1991) also observed in most cases, that one of the random effects commonly overshadows the other. If the strength of μ_i exceeds that of v_i , then the risk estimated will display a spatial structure and if the strength of v_i supersedes that of μ_i conversely, it means the implication will be to compress the estimated means towards the total mean.

Also, Besag *et al.,* (1991) assumed that the terms u and v were independent with the following priors:

$$p(v \mid \tau) \propto \tau^{-n/2} exp\left\{\frac{1}{2\tau}\sum_{i=1}^{n} v_i^{\,2}\right\} \tag{2.34}$$

and

$$p(u \mid k) \propto k^{-n/2} exp\left\{-\frac{1}{2k}\sum_{i}^{n}\sum_{j\in N(i)}(u_i - u_j)^2\right\} \tag{2.35}$$

Fundamentally, equation (2.34) means that v, the spatially unstructured component consist a Gaussian process white noise with a variance τ that is unknown, and equation (2.35) denotes that the spatially structured component term u, also has a known Gaussian Markov random field (GMRF) process with variance k, with n given as the number of areas under study and $N_{(i)}$ denotes the set of neighbours of area *i*. This neighbourhood can then be defined in terms of Euclidean distance of the centroids of the areas, if the two regions share a boundary or a combination of these two. Besag *et al.,* (1991) defined their neighbourhood based on shared border.

This study was concentrated upon by Bernardinelli *et al.,* (1995) and reached a decision that the standard deviation of the conditional distribution of the spatially correlated random effects would be 0.7 times the standard deviation of the spatially uncorrelated random effects. Though, this assumption is still made open for in-depth study and discussion. If given that the prior distributions of the precision parameters for the two random effects expressed in the convolution model are assumed to be non-informative, it then means only the summation of these two random effects will be classifiable and not as separate components.

This suggests that when the rest of the conditional distributions of each u_i are given, then we have

$$\left(u_i \mid u_{-i}\right) \sim N\left(\frac{\sum_{j \epsilon N_i} u_j}{d_i}, \frac{k}{d_i}\right) \tag{2.36}$$

with

$$E(u_i - u_{-i}) = \frac{\sum_{j \in N(i)} u_j}{d_i} \tag{2.37}$$

and

$$\text{Var}(u_i \mid u_{-i}) = \frac{k}{d_i} \tag{2.38}$$

where *di* denotes the amount of neighbours in region *i*. This distribution of *u*, which is conditional, is known as the intrinsic conditional autoregressive (ICAR) prior distribution. Besag *et al*., (1991) investigated the posterior distribution by using the Gibbs sampler, a part of an MCMC algorithm.

2.10 Spatial Random Effects

Modelling the count of people with a particular disease within a set of areas $i = 1, 2, \ldots, n$ can be considered from a hierarchical perspective, where we model the disease count data as Poisson. Random effects are then introduced to make up for spatial dependence unexplained by this observed data. These random effects also constitute spatial heterogeneity in addition to spatial correlation in the count data on the disease. In spatially indexed data set, spatial random effects connected to each area or location may be used, letting for the modelling of a fundamental spatial dependence structure (Osei, 2010).

Considering the modelling framework in the Bayesian approach allows for adjustment of the uncertainty parameters by allocating prior distributions to the parameters.

The spatial link structures are then integrated in a Bayesian situation when appropriate prior distributions are assumed for the spatial random effects. There is a need to assume a prior distribution for the spatial effects so as to impose a spatial dependency structure which takes into account, the neighbourhood arrangement of the study region where the count of people with a particular disease is conducted.

2.10.1 Additive and Multiplicative Random Effects

These types of random effects terms are presented to account for unnoticed or ignored spatial heterogeneity inherent in the count data set. Depending on how we choose to model the spatial count data, these random effects can either be additive or multiplicative. It is generally well known that count data show over-dispersion when linked to the Poisson distribution. It has been established that the unobserved heterogeneity normally seen to be the basis of over-dispersion in count data models has some predictable consequences for the probability structures of such models (Mullahy, 1997).

Disease counts data which are generally assumed to follow the Poisson distribution has a mean parameter λ which can be modeled in two ways:

(i) Adding an unobserved random variable ε.

(ii) Multiplying by an unobserved random variable ε.

The first case is called the additive random effects, which conditional autoregressive models use when accounting for spatial heterogeneity in the count data. On the other hand, the multivariate Poisson gamma uses the second case which is the multiplicative random effects.

2.11 Spatial Heterogeneity

Two authors, Fotheringham *et al.,* (2000) and Lesage (1999) also raised an additional subject and it centers on spatial heterogeneity. It could either be in terms of regression relationships or in terms of heteroscedasticity observed in a spatially unstructured error term as discussed above. Reasons could also be problems of identifiability in splitting spatial dependence from spatial heterogeneity (de Graaff *et al.*, 2001; Anselin, 2001). In disease mapping applications, the word excess heterogeneity is frequently applied to spatially unstructured errors, for Poisson over-dispersion in the log link for count outcomes.

In analysing small area disease data in epidemiology, the central purpose repeatedly, is estimating the true nature of relative risk in the presence of over-dispersion in the observed event counts and spatially correlated errors as a result of lost predictors. The assessment of relative risks by some standard methods built on the Poisson distribution gives that a disease risk is often the same over areas, and also over the individuals inside that area. Individuals may not be the same within the areas, and the disease risks differ between the areas, with the intention that the observed counts display a larger unevenness than what Poisson shows.

This disparity can be modeled by displaying the area relative risks in terms of one or more random effects (Congdon, 2003). Effects of such may be spatially unstructured and which tends to white noise in time series, and these are sometimes called extra heterogeneity (Best *et al.*, 1999). While heteroscedasticity can well be relevant, a known alternative viewpoint on spatial heterogeneity pays attention to the regression parameters themselves.

Both linear and general linear models characteristically adopt that the structure of the model remains homogenous over the study region without any local variations in the parameter estimates.

In several applications, the effects of regression are likely not to be constant over the area of application (Casetti, 1992). An approach for the enhancement of relative risk estimation is to use smoothing tools on Standard Mortality ratio (SMR) to reduce the noise. To overcome this, the method of conditional autoregressive model (CAR) is used.

The models (CAR) are usually used to define the spatial difference or variation of quantities of importance in the form of aggregates over subregions. These CAR models have been employed to analyse data in various areas like demography, economy, epidemiology and geography. Overall accounts of CAR models are given as a class of Gaussian Markov random fields (GMRF) in Cressie (1993), Banerjee *et al.,* (2004) and Rue and Held (2005).

The models have been broadly used in spatial statistics to model observed data (Cressie and Chan, 1989; Richardson *et al.,* 1992; Bell and Broemeling, 2000; Militino *et al.,* 2004; Cressie *et al.,* 2005), as well as (unobserved) latent variables and spatially varying random effects (Clayton and Kaldor, 1987; Sun *et al.,* 1999; Pettitt *et al.,* 2002; Banerjee *et al.,* 2004).

This model was first introduced by Besag (1974). The hierarchical version of disease-mapping models that is based on CAR was studied by Besag *et al.,* (1991). Proper multivariate conditional autoregressive models were used by Gelfand and Vounatsou (2003). Venkatesan *et al.,* (2008 and 2010) studied HIV and tuberculosis patterns for India using CAR. Normally, a CAR model offers a facet in defining the spatial autocorrelation structure that exceeds distance based functions and combines the idea of spatial neighbours. Besag *et al.,* (1991) allocated weights based on whether a pair of regions shared a borderline or not; if the regions share a boundary, the weight is allocated 1, otherwise it is allocated 0. Other weighting possibilities were discussed in Best *et al.* (1999).

The overall objective of these spatial models is to reveal and measure spatial relations present between the data, particularly to measure how quantities of interest differ with independent variables and to identify bands of 'hot spots' (Oliveira, 2012). It is also common to assume that observations we make at locations close to each other have a tendency to have interrelated values. This conditional autoregressive (CAR) model is therefore a standard choice for analysing these lattice data. Another alternative is to use CAR to directly model the covariance structure of disturbance terms in a linear regression model (Anselin, 1988).

A model like the CAR have been broadly used in spatial statistics in modelling observed data (Cressie and Chan 1989; Richardson *et al.,* 1992; Bell and Broemeling 2000; Militino *et al.,* 2004; Cressie *et al.,* 2005), also to model (unobserved) latent variables and spatially changing random effects (Clayton and Kaldor, 1987; Sun *et al.*, 1999; Pettitt *et al.*, 2002; see Banerjee *et al.*, 2004). The CAR model is also used extensively in Bayesian models as prior distribution for random spatial effects (Renato *et al.,* 2009). Spatial data in Bayesian analyses frequently use the conditionally autoregressive (CAR) prior promoted for mapping disease by Besag *et al.*, (1991). The central idea for a conditional autoregressive model is that the probabilities estimated at any given location, say *i*, are conditional on the level of neighbouring values.

Several writers have also suggested other designs for the convolution (CAR) model. Particularly, Leroux *et al.*, (1999) as different from Besag *et al.*, (1991) design of a random intercept divided into two parts, only one random intercept was used by the authors and its variance covariance matrix was divided into spatial and non-spatial components, with a parameter directing the spatial dependency. For some authors with different proposals, check MacNab and Dean (2000) for a parametric bootstrap method, and also see Green and Richardson (2002) for method on their hidden Markov field.

Despite all these substitute models, the model suggested by Besag *et al.,* (1991) still continue to enjoy greater application as a result of its close fit with the usual MCMC operations and also because of a broad assortment of freely accessible software, e.g WinBUGS (Spiegelhalter *et al.*, 2007) with R statistical software (R Development, 2008) effecting it.

In the case of CAR, the random effect conditional distribution in an area when all others are given is simply given as the weighted average of all the other random effects. The popular Besag *et al.,* (1991) allocated the weights centered on whether an area shared a boundary or not as earlier stated in this review proceedings; if the locations share a boundary, the weight is allocated 1, otherwise it is allotted the value 0. Best *et al.,* (1999) also discussed extra weighting potentials.

The weighting alternatives debated thus far are expected to be stable and fixed during modelling. Another approach was taken by Lu *et al.,* (2007). This weighting is done by giving the estimates from the data itself. The CAR is known to have a computational benefit against the multivariate Gaussian for the reason that the variance component in multivariate Gaussian needs matrix inversion in its estimation, for every update when performing an algorithm giving rise to more computational problem and which in CAR is not essential. Below are some of the models used for spatial dependency in disease mapping.

2.11.1 Intrinsic CAR (Besag) Model

The simplest of the CAR prior is the intrinsic model suggested by Besag *et al.,* (1991), which is given as

$$\phi_i|\phi_{-i} \sim N\left(\frac{\rho\sum_{j=1}^{n}\omega_i\phi_j}{\sum_{j=1}^{n}\omega_{ij}}, \frac{1}{\tau\left(\sum_{j=1}^{n}\omega_{ij}\right)}\right), \qquad i=1,\ldots,n \tag{2.39}$$

where the parameter, τ, is the conditional precision. The precision is comparative to the number of neighbouring units, while the conditional expectation of ϕ_i, is the mean of the random effects in the neighbouring areal units. The precision formulation here is functional, because you would expect the precision to be higher when you have more neighbouring areas and therefore more information to estimate the value of ϕ_i. These groups of conditional distributions agree to the multivariate normal distribution, with a zero vector mean. The improper precision matrix is given by $Q = \tau\left(diag\left(W1\right) - W\right)$, with $W1$ as a vector comprising the number of neighbors for each of the areal unit.

One limitation of this model is the lack of a parameter to regulate the strength of the spatial autocorrelation; if you multiplied ϕ by 10, then the precision, τ would decrease, but the spatial structure does not change. This implies that the intrinsic model is only practical in cases where the spatial autocorrelation in the data is strong; it is not practical for cases where there is weak or moderate spatial autocorrelation across the study region because the model would tend to produce an overly smooth estimated risk surface in these cases.

This formulation of the precision will make sense if strong spatial autocorrelation is present, because an increased number of neighbours mean that more information is available to estimate the random effect value. However, in cases where weaker spatial autocorrelation is present, this formulation is less practical, because by increasing the number of neighbours would not certainly lead to a huge increase in the amount of information available to estimate the random effect value.

2.11.2 The BYM Model

The assumption of the Besag model adopts that a spatially structured component cannot take the limiting form that allows for no spatially structured variability. Hence, unstructured random error or pure over-dispersion within area *i*, will be modelled as spatial correlation, giving confusing parameter estimates (Breslow *et al.,* 1998). To address this issue, the Besag-York-Mollie (BYM) model (Besag *et al.*, 1999) splits the regional spatial effect **b** into a sum of an unstructured and a structured spatial component, so that **b = v + u**.

Here, $v \sim \aleph(0, \tau_v^{-1}\mathbf{I})$ accounts for pure over-dispersion, while $u \sim \aleph(0, \tau_u^{-1}\boldsymbol{Q}^{-})$ is the Besag model whereby $\boldsymbol{Q}^{-}$represents the generalized inverse of$\boldsymbol{Q}$. The subsequent covariance matrix of **b** is

$$Var(b|\tau_u, \tau_v) = \tau_v^{-1}\mathbf{I} + \tau_{\boldsymbol{u}}^{-\mathbf{1}}\boldsymbol{Q}^{-}.$$

2.11.3 Leroux CAR Model

In the BYM model, the unstructured and structured components cannot be seen independently from each other, and are thus not identifiable (MacNab, 2011). Leroux *et al.*, (2000) proposed an alternative model formulation to make the compromise between unstructured and structured variation more explicit. The concern of accounting for the chance of weaker spatial autocorrelation was addressed by Leroux *et al.*, (1999), who proposed the following CAR model:

$$\phi_i|\phi_{-i} \sim N\left(\frac{\rho\sum_{j=1}^{n}\omega_{ii}\phi_j}{\sum_{j=1}^{n}\omega_{ij}\rho+1-\rho}, \frac{1}{\tau\left(\sum_{j=1}^{n}\omega_{ij}\rho+1-\rho\right)}\right), \quad i=1,\ldots,n \qquad (2.40)$$

Here, ρ controls for the degree of spatial autocorrelation currently in the data. For a value of ρ = 1 corresponds to the intrinsic model (2.39), while ρ = 0 corresponds to a completely spatially smooth model with a constant mean, 0, and precision, τ. This increased flexibility can therefore enable the random effects to model a wider range of spatial autocorrelation than the intrinsic approach.

2.11.4 Lee CAR Model

The intrinsic and Leroux models are both globally smooth; that is they assume a constant level of spatial smoothness across the whole study region with the partial correlation between (ϕ_i , ϕ_j) conditional on the remaining random effects ϕ_{-ij} given by

$$Corr\left[\phi_i, \phi_j\right] = \frac{\rho\omega_{ij}}{\sqrt{\left(\rho\sum_{k=1}^{n}\omega_{ik} + 1 - \rho\right)\left(\rho\sum_{l=1}^{n}\omega_{jl} + 1 - \rho\right)}} \qquad (2.41)$$

For the Leroux CAR, a parameter ρ close to 1 will give a strong spatial autocorrelation between all pairs of adjacent areas for which $\omega_{ij} = 1$,while if ρ is close to 0 then there will be lower spatial autocorrelation across the study region. Thus ρ moderates the level of spatial smoothness across the region globally. This may not be true in practice, because you may expect different levels of spatial autocorrelation in different areas of the study region. This was being addressed by Lee *et al.,* (2014), where he suggested a localized conditional autoregressive model which offers more flexibility in the way the random effects are modeled by permitting for discontinuities in the spatial autocorrelation surface.

Here, elements of the neighbourhood matrix relating to adjacent areas, $\{w_{ij}|i \sim j\}$, are treated as binary random quantities which are no longer fixed at 1; if ω_{ij} is estimated as 0 for neighbouring areal units i and j, then that corresponds to a boundary between the areal units as (ϕ_i , ϕ_j) are conditionally independent given the random effects. The matrix W is defined as a set of edges, and under this terminology $w_{ij} = 0|i \neq j$; means that an edge has been removed.

2.11.5 The Dean Model

Dean *et al*.(2001) proposed a reparameterisation of the BYM model where

$$b = \frac{1}{\sqrt{\tau_b}}\left(\sqrt{1-\emptyset}v + \sqrt{\emptyset}u\right), \qquad (2.42)$$

having covariance structure

$$\text{Var}\left(b|_{\tau b,\emptyset}\right) = \tau_b^{-1}\left((1-\emptyset)\mathbf{I} + \phi Q^{-1}\right) \qquad (2.43)$$

Equation (2.42) is a reparameterisation of the original BYM model, where $\tau_u^{-1} = \tau_b^{-1}\phi$ and $\tau_v^{-1} = \tau_b^{-1}(1-\phi)$ (MacNab. 2011).

The additive decomposition of the variance is then on the log relative risk scale. This is in contrast to the Leroux model (2.40), where the precision matrix of b resulted as a weighted average of the precision matrices of the unstructured and structured spatial components. As a consequence, the additive decomposition of variance in the Leroux model happens on the log relative risk scale, conditional on $b_j j \in \delta i$ (Leroux *et al*., 2000).

2.12 Bayesian Ranking Methods in Disease Mapping.

In disease mapping, early attention to emerging hotspots, that is, areas with high risks which are surrounded by areas with much lower risks, before they become extreme, is vital in decision-making related to health surveillance. Such decision making processes may refer to optimal allocation of resources for health prevention, or to decisions reflecting mobility of a society or other environmental controls.

A typical approach for detection of disease hotspots through a hypothesis testing framework uses the scan statistic (Kulldorff and Nagarwalla, 1995; Kulldorff *et al.*, 1998), which aims at detecting the location and size of hotspots without any defined assumptions about these values. It is very necessary to estimate and rank various local elevations in risk across a map. For rare diseases like tuberculosis, the observed disease count may exhibit extra Poisson variation. Hence, the standardized mortality ratios (SMRs), a basic investigative tool for epidemiologists, may be highly variable. Consequently, in maps of SMRs, the most variable values, arising normally from low population areas, tend to be highlighted, hiding the true underlying pattern of disease risk.

To address the issue of such over-dispersion, the field of disease mapping has increased in the last decade with a range of valuation approaches and spatial models for latent levels of the hierarchy of the model. In particular, there have been many developments related to Bayesian hierarchical models, which permit an area risk to borrow strength and information from the neighboring areas where the disease risks are similar. These models have certainly become standard tools for mapping disease rates (see Besag *et al.*, 1991; Clayton and Bernardinelli, 1992; Clayton *et al.*, 1993; Lawson *et al.*, 2000; MacNab *et al.*, 2004; Best *et al.*, 2005, for example) in order to recognize global hotspots and trends in the risk surface across the map.

Identification of local or developing hotspots has received less attention. It is uncertain whether and what sorts of smoothing techniques offer advantages for identifying isolated hotspots, over basic estimates such as raw rates.

Here, we still maintain the emphasis on Bayesian hierarchical conditional autoregressive (CAR) models, developed by Besag *et al.*, (1991); Clayton and Bernardinelli (1992); Clayton *et al.*, (1993). The CAR model and its extensions have become commonplace in epidemiological studies and have been shown to be flexible and robust (Lawson *et al.*, 2000). Best *et al.*, (2005) demonstrated the merits of the CAR model when compared to other modern models together with a multivariate normal geostatistical model with an exponential covariance, a spatial mixture model type, a partition type model and a gamma type moving average model.

While CAR models were not intended to identify isolated hotspots of isolated hotspots, but have been used broadly for identifying extreme risks. The most natural measure of isolation is the difference between the risk or rank of a potential hotspot and the corresponding quantity for its neighbors. Ranking methods play a valuable role in drawing attention to elevated regions.

Laird and Louis (1989) showed that ranking of empirical Bayes estimators can be more accurate than that of conventional maximum likelihood estimators. Shen and Louis (1998) investigated ranking procedures using loss functions of squared error operating on the difference existing in the estimated and exact ranks. Though we focus on disease mapping, we note that methods for ranking isolation measures may be broadly useful in many other contexts, particularly sociological, for ranking political or racial isolation, or ecological, for diversity studies.

2.13 Some Empirical Reviews of Bayesian Spatial Modelling Techniques with Application to Tuberculosis Data.

Research to appraise the spatial distribution of TB and to identify high-risk areas is limited especially in developing countries (Bishai *et al.*, 1998; Verver *et al.*, 2004; Vieira *et al.*, 2008). Since it is known that control efforts are best intended when areas of high prevalence are known, it is also important to know areas where rates are unusually high given the underlying risk factors.

Spatial analysis techniques have in recent times been added as an influential tool in public health, because of its ability of envisaging the distribution of disease, even with scarce data (Best *et al.*, 2005), and disease mapping risk factors observed at population level (Bailey, 2001; Bailey *et al.*, 2005). Some current write-ups have highlighted the usefulness of Geographical Information Systems technology in Public Health research, but its use being inadequate arising from the worth of the address information obtained, mainly those obtainable in monotonous information systems.

Some researchers have been able to identify socioeconomic status (SES) as an essential cause of the observed social inequalities in health outcomes (House *et al.*, (1990); Williams (1990); Link *et al.*, (1995)) and in particular of racial differences in health (William 1997).

A vast review of accessible literature shows that not much research have been done in building Bayesian spatial models for TB as a single outcome health case in South Africa. However, some authors and articles have established some spatial (Bayesian) models which have been applied to many countries and which depend on the nature of the epidemic.

Tiwari *et al.,* (2010) investigated tuberculosis cluster in Dehradun city of India using Geographical Information System (GIS) and spatial scan statistics. Their objective was to examine the occurrence of statistically significant geographical assembles of tuberculosis in that city so as to detect their nearest locations.The spatial scan statistic saTScan v6.1 used is centered on the likelihood ratio test (LRT) and detecting of high/low clusters was carried out under Poisson probability model assumption. All the geographical and cartographic outputs were presented by the ArcGIS 9 application software. Significant enormous clusters were recognized in seven wards of the Dehradun Municipal area.

Bastida *et al.,* (2012) also identified the spatial and temporal distribution of TB for a four year period in a Mexico state, by the use of GIS and SCAN statistics program.

They identified nine major clusters ($P< 0.05$) by spatial-time analysis. Their decision was that, TB prevalence in the state of Mexico is not randomly distributed but systematic and is concerted in areas close to Mexico City. A strong relationship between the strains of TB clusters and the distance to the center of urban zones was also found.

Maciel *et al.,* (2010) investigated spatial arrays of the occurrence of pulmonary TB and its association with socio-economic status in Vitoria, Brazil. Spatial Poisson models were fit to examine the association between socio-economic rank and TB incidence alongside with Anselins local indicators of spatial clustering statistics (LISA), smoothed Bayesian estimates and model-predicted incidence rates to show the spatial patterns of disease incidence.

The index of quality of urban municipality (IQU) was calculated for each neighborhood to measure for socio-economic status. The IQU, developed by the Instituto Pólis (São Paulo, SP, Brazil), is a simple arithmetic mean with 11 variables and a range between 0 and 1, with 1 being the highest level.

They used a supervised backward elimination method to select model variables using both the methods of Akaike's information criterion (AIC) and Bayesian information criterion (BIC) statistics as criteria (lesser values considered improved fit) with the full model written:

$$ln\ (count)\ =\ ln\ (pop)\ +b_0+b_1.IQU\ +b_2.(IQU-\mu)^2+b_3.(IQU-\mu)^3+b_4.Lat\ +b_5.Long$$
..(2.44)

Each beta estimate is interpreted as the increase in log-incidence for a one unit increase in its respective variable, where b_1, b_2 and b_3 are the beta parameters for the IQU (linear, quadratic, cubic), and b_4 and b_5 are beta parameters for latitude and longitude, respectively. The model was fit using PROC GLIMMIX in SAS 9.2 (SAS Institute, Cary, NC, USA). The method of Moran's *I* showed solid spatial autocorrelation between incidence rates and four groups of high occurrence were recognized by LISA. TB incidence alongside socioeconomic status had a significant curvilinear relationship.

Wayner *et al.,* (2007) investigated the consequence of socio-economic deficiency and some transmission-related pointers of tuberculosis (TB) cases at small area level, to debate the potential of each indicator in aiming locations for developing protective action. The authors employed four full hierarchical Bayesian models differently to estimate the relative risk of the incidence of TB through Markov chain Monte Carlo. A generalized linear mixed model (GLMM) was used to smooth out the unevenness in the observed disease rates and to estimate the relationship between TB incidence rate, over an observed period and some chosen covariates. Considering the Bayesian context though, the observed number of cases in $y=(y_1,\ldots,y_n)$ the *n* regions is one recognition of Poisson random variables, $(y_1,\ldots,y_n)$, with means, $\mu=(\mu_1,\ldots,\mu_n)$.

This value was stated as the amount of expected cases ε_i multiplied by the relative risk of each area: $\mu_i = \varepsilon_i \times \lambda_i$, or, in logarithm form, $log\,\mu_i = log\,\varepsilon_i + log\,\lambda_i$. The relative risk λ_i is defined as a function of the *k* covariates x_k which explains the differences in disease rate. A non-informative Gaussian priors with zero mean and precision equal to 1×10^{-5} was assumed for β.

Still in that framework, v_i is a non-spatially structured random effect, normally assumed to be independent Gaussian, with zero mean and variance, σ_v^2 usually incorporated into the models to explain for extra-poisson variation, because of non-measured significant covariates. The spatially structured random effects parameter – $\theta = (\theta_1, \ldots, \theta_n)$ – accounts for the spatial dependence, with the prior distribution taken as a conditional intrinsic Gaussian autoregressive model, where the average value for θ_i is a weighted mean of the neighbouring random effects and the variance, $\sigma_{\theta 2}$, directs the strength of this local spatial reliance.

$$P(\theta_i | \theta_{j \neq i}) \sim N\left[\frac{\sum_{j \neq i} \omega_{ij} \theta_j}{\sum_{j \neq i} \omega_{ij}}, \frac{\sigma^2}{\sum_{j \neq i} \omega_{ij}}\right] \qquad (2.45)$$

Guy *et al.*, (2013) examined the spatial distribution, alongside the social and economic correlates of tuberculosis in Brazil during the period 2002 and 2009 using municipality-level age/sex-standardized tuberculosis notification data. The disease rates were strongly spatially autocorrelated, being remarkably high in municipal regions on the eastern seaboard and in the west of the country.

Non-spatial ecological regression analyses showed higher disease rates related with urbanicity, population density, poor economic conditions, household crowding, non-white population and worse health and healthcare indicators.

All these relationships remained in spatial conditional autoregressive models, though the outcome of poverty seemed to be partially confounded by urbanicity, race and spatial autocorrelation, and partially interceded by household gathering. Their study emphasized both the multiple associations between socioeconomic factors and tuberculosis in Brazil, and the significance of accounting for spatial factors in analysing socioeconomic factors of tuberculosis.

Roza *et al.,* (2012) studied to investigate the urban spatial and temporal distribution of tuberculosis (TB) in Ribeirão Preto, State of São Paulo, southeast Brazil, during the years 2006 to 2009 and to assess its association with causes of social vulnerability such as income and education level. Social vulnerability measures were acquired from the SEADE Foundation, and information about the number of inhabitants, education and income of the households were accessed from Brazilian Institute of Geography and Statistics. The Statistical analyses were shown by a Bayesian regression model that assumed a Poisson distribution for the observed new cases of TB in each area. For the spatial covariance structure, a conditional autoregressive model was used. The Bayesian model established the existence of spatial heterogeneity of TB distribution in Ribeirão Preto, identifying higher risk areas and also the effects of social vulnerability on the disease. The authors established that the high incidence of TB was related with the measures of income, education and social vulnerability.

Sampurna *et al.*, (2014) studied the spatial and temporal variations in TB incidence in Nepal. They established models for TB incidence by gender, year and location using linear regression of log-transformed incidence rates. Aside a comparatively small amount of outliers, a good fit was produced by their models as showed by residual plots and estimates from the r-squared statistic. They observed that there were trends of increase in TB for current years among sexes. The additive log-linear regression model that was used to model TB incidence rates with normally distributed errors that differ by location and year was defined as:

$$ln\left(n_{ij}/P_i\right) = y_{ij} = \mu + \alpha_i + \mathrm{b}_j \qquad (2.46)$$

In their model, P_i was the corresponding population at risk in 1000s and the terms α_i and β_j signifies the super-district and gender-year effects that sum to be zero so that μ is a constant capturing the total occurrence. The model fit was then evaluated by the linearity in the plot of deviance residuals versus the normal quantiles. Also the model gave adjusted prevalence rates for each cause of interest, acquired by replacing the parameters consistent to the other factors by constants chosen to make sure the total expected number of cases equals the observed number.

A spatial modelling method is basically concerned with these three main concerns and they are estimation and inference of parameters, comparison and specification of the model, and also prediction, known as kriging. In the disease spatial epidemiology investigation, there exists a concern in discovering the spatial patterns of disease in exact areas and locations and to investigate whether these patterns have any spatial dependence.

Kriging is one of the interpolation methods which are used to locate the spatial dependence and to extrapolate the sites of cases from unmeasured sites. A few reviews in this dimension will also be considered in this section.

Spatial variations of Pulmonary TB occurrence and its association with socio-economic and geographical factors in China were studied by Xin-Xu Li *et al.*, (2014). The authors performed an evaluation using ArcGIS to choose which kriging and cokriging devices along with diverse combinations of types of detrending, semivariogram models, anisotropy and covariables (socio-economic and geographical factors) could properly construct spatial distribution surface of PTB incidence using statistic data. They found that the global cokriging with socio-economic and geographical factors as covariables confirmed to be the best statistical approach for precisely estimating spatial distribution surface of PTB prevalence.The forecast constant surface then exemplified the spatial variations of PTB incidence that were co-impacted by socio-economic and geographical factors.

Srinivasan *et al.*, (2013) used a Bayesian kriging method to investigate the spatial arrangement of tuberculosis within a Chennai ward population in India to gain an insight into the disease spread and also, to infer the locality from the unmeasured locations. Moran's I technique was one of the approaches used to see the autocorrelation of the disease based on the region which is one of the oldest pointers of spatial autocorrelation (Moran, 1950). It is used to measure the strength of spatial autocorrelation in a map. It uses the cross-products deviances from the average and it is calculated for n observations on a variable x at regions i,j as: Ordinary and Bayesian kriging methods were employed in their geo-statistical predictions. The latter method was found to bring additional elasticity to the standard prediction framework.

More explicitly, prediction in Bayesian develops from the posterior predictive distribution which mixes above the posterior distribution of all model parameters, that is:

Orinary kriging,

$$Y(S) = \mu + \epsilon(S) \tag{2.47}$$

and Bayesian kriging,

$$P\big(Z(S_o)\big|Z, X, x(S_o)\big) = \int P(Z(S_o), \beta, \alpha, \eta, \theta | Z, X, x(S_o) d\beta dx d\eta d\theta$$

$$= \int p\big(Z\big(S_0\big) \mid Z, X, x\big(S_0\big), \beta, \alpha, \eta, \theta\big)\, p\big(\beta, \alpha, \eta, \theta \mid Z, X\big) d\beta d\alpha d_{\eta} d\theta \tag{2.48}$$

Venkatesan *et al.,* (2010) modeled the spatial variogram of tuberculosis for Chennai ward in India to describe the spatial dependence of TB in that ward. A variogram is used to evaluate whether disease characteristics of the cases are random or clustered. In their studies, they used three models namely, the exponential, the spherical and the Gaussian model as theoretical models and compared it with their variogram model, to mathematically describe the shape of the disease. In their comparisons using AIC, AICC and BIC, the spherical model fit data better.

Bayesian methods of disease epidemiology also make attempts to present a good posterior relative risk of the disease cases as against the raw risk patterns which could be misleading. Bayesian Conditional Autoregressive models, (CAR), are suitable for smoothing disease relative risk estimates based on neighbourhood structure, which provides some shrinkage of the raw relative risk estimate of the pattern of underlying risk factor than that produced by raw estimates.

Venkatesan *et al.*, (2012) constructed a Bayesian spatial CAR to study TB patterns in India to smoothen the relative risk of the disease. He compared the Bayesian log-normal relative risk with the Bayesian CAR, which he proposed. The Deviance Information Criterion (DIC) approach was used in comparing the two models and it was found that Bayesian CAR showed less deviance and which showed that CAR model is best suitable for mapping of the disease.

Also, mapping of disease has received many applications in epidemiology and public health. The fundamental model normally used in disease charting is the Besag, York and Mollie model (BYM), which usually combines two random effects, one which is spatially structured and the other random effect which is spatially unstructured. Assumption of normality on the uncorrelated random effect in models is common and which is mainly because of its computational simplicity. At times this assumption is improper because some random effects can be platykurtic, leptokurtic or skewed; deviating from this general normality assumption (Box and Tiao, 1973). It is therefore necessary to examine a more symmetric, flexible and robust distribution for the spatially unstructured random effect by considering another related type of Gaussian distribution in the disease mapping problem.

Rindra *et al.*, (2010) studied to explore the pattern of tuberculosis for Antananarivo, Madagascar and the relationship that exists between the spatial variation of TB risk and national control program pointers for all neighbourhoods using a Bayesian method. Mixture of a Bayesian methodology and a generalized linear mixed model (GLMM) was used to measure spatial heterogeneity in the TB standardized incidence ratio (SIR) and to examine relationships between the three year average TB incidence rates and some variables.

Contained in the Bayesian context, the observed numbers of new cases $y_i = (y_1, ..., y_n)$ in the n neighbourhoods were treated as non-independent Poisson random variables with means $\mu_i = (\mu_1, ..., \mu_n)$, where each μ_i is given by

$$\mu_i = E_i \times \lambda_i,$$

or, in the logarithmic form,

$$log(\mu_i) = log(E_i) + log(\lambda_i). \qquad (2.49)$$

The SIR λ_i is a known function of the explanatory variables X_{ki} that account for differences and spatial heterogeneity in the disease rate:

$$\lambda_i = exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \theta_i + \nu_i) \qquad (2.50)$$

with $X_{ki} = X_{ki} / SD_k$, where SD_k is the standard deviation (of all neighbourhoods) of each variable. There result discovered that there were clusters of high TB risk areas and the distribution of TB was found to be related primarily with the number of patients lost to follow-up and the number of households with more than one disease case.

Also, in most count data, the average and variance are frequently connected and can be estimated by the use of a single parameter, as in the Poisson model, which is the most commonly used model for disease mapping data analysis.

It is observed that much has not been done in the literatures of modelling the patterns and incidences of tuberculosis as a single disease with over-dispersion and as a count data. Previous studies outlined in this thesis shows that tuberculosis modelling was done by assuming a Poisson type of regression, disregarding the existence of over-dispersion and modelling it in the data.

Modelling of count data in the presence of over-dispersion has been studied for other diseases that constitute counts. A rare disease type like tuberculosis that generates count data have not been widely studied and applied in Eastern Cape Province of South Africa.

For instance, Mohammadreza *et al.,* (2014) in their paper applied the generalized linear model to model geographical variation in esophageal cancer incidence data in the Caspian region of Iran. Their data constitute a multifaceted and hierarchical structure that make them appropriate for hierarchical analysis by the Bayesian techniques, but which caution was essential to deal with problems arising from counts of actions detected in small geographical areas when over-dispersion and residual spatial autocorrelation are present. The authors used the Poisson, generalized Poisson and negative binomial, and three different autocorrelation structures. The confirmation from applying the modelling method suggests that the modelling approaches from the use of the generalized Poisson and negative binomial with spatial autocorrelation worked well and provided a robust basis for inference.

Another very commonly used modelling approach that deals with over-dispersion and count data is called the zero-inflated models - a type of mixture models and approach employed in the analysis of count data when there exist over-dispersion.

One recurrent expression of over-dispersion is that the frequency of zero counts exceeds the expected for the Poisson distribution and this is interesting because zero counts normally have special prominence. For data that consist of zero values, zero inflated Poisson (ZIP) is the suitable model to substitute the Poisson models.

Musenge *et al.,* (2013) used data from Agincourt located in the rural northeast of South Africa, collected longitudinally during the years 2000 and 2005.Agincourt child HIV deaths data had several problems that made it difficult to use standard statistical procedures: over dispersion, caused by unobserved heterogeneity or temporal correlation, large data, excess of zeros, and household spatial random effects. In their paper, they applied two zero inflated models adjusting for household spatiotemporal random effects using Bayesian inference and performed exploratory analyses looking for risk influences for childhood HIV/TB death rates using spatial modelling. The modelling procedure adjusts for the household level spatial random effects for child HIV/TB mortality. The method of Bayesian zero inflated spatiotemporal models were able to spot hidden patterns within the data. Geo-additive spatiotemporal zero inflated Poisson (ZIP) as well as Negative Binomial (ZINB) models were employed in the analysis. The ZIP and ZINB models have many parameters and are hierarchical (multilevel), thus we resort to full Bayesian inference with the computationally efficient MCMC techniques. Their main finding revealed that those maternal orphans were almost thrice at bigger risk of HIV/TB mortality matched with those with living mothers. Also, a risk analysis which amends for person, place and time enables policy makers to use estimates and maps for interventions.

Mohd *et al.,* (2013) in their study also modeled a joint disease, HIV-TB mortality for the period between 2001- 2009 for Kelantan state in Malaysia.

The study modeled this scenario using the Zero Inflated Negative Binomial model. In conclusion, their study recognized a high complete mortality rate among patients hospitalized with HIV and TB. However, their study posit that for ZIP and ZINB models to give good statistical fit but do not distinguish the underlying mortality process, then numerous stimulating questions may arise. First could be that, what are the consequences of applying these models? Secondly, what substitute model should be used instead? The suitable and handy answer for these two questions, are if the only objectives should consist of achieving the best statistical fit other than the zero inflated model.

2.14 Conclusion

Our evaluation of Bayesian concepts in spatial statistics only traces on a sequence of common subjects through major and common issues in spatial statistics. The principal factor to all methods (for all Bayesian analyses) is the improvement and application of a fully defined probability model from which posterior inference follows. According to Box and Tiao (1973, pp. 9-10), the suitability of the model can always be tested, but fully Bayesian (posterior-based) inference focuses criticism and disapproval on the correctness of the probabilistic structure of the model, rather than the mode of inference, correctness of approximation, or suitability of the particular choice of estimator.

As earlier stated in this literature review that Bayesian analyses is basically concerned with model specification, model comparison and prediction, this thesis therefore focuses on developing and specifying alternative models and distributions to the convolution model (BYM;

Besag *et al.,* 1991), by specifying alternative models and comparing them for robustness and flexibility based on this tripod operations of Bayesian analyses.

From our reviews, modelling responses in count data therefore require that adjustments be employed on the BYM model, by evolving other probability densities that can model over-dispersion that is common with count data. Mixture models from Poisson as alternatives to the Poisson density to model over-dispersion have been suggested. Among those suggested are the negative binomial (NB) models, which is a good and promising model to adjust if over-dispersion is found. However, some authors have pointed out that the negative binomial model itself can be over-dispersed, because it can converge to a Poisson density at some certain amount of the dispersion parameter (Imoto, 2014; Xinyan *et al.,* 2017).

The contrary case is that of under-dispersion, where the variance is smaller than the mean. Although literature contains more examples of over-dispersion, under-dispersion is also common. Rare events, for instance, generate under-dispersed counts. Examples include the number of strike outbreaks in the UK coal mining industry during successive periods between 1948 and 1959 and the number of eggs per nest for a species of bird (Blincoe *et al*., 2000).

In such cases, neither the Poisson nor the negative binomial distributions provide adequate approximations. This thesis would therefore explore some Poisson mixtures to deal with this gap by comparing some of these alternatives with existing ones.

Another task in spatial analyses is the problem of spatial autocorrelation. A way of dealing with over-dispersion is also by means of random effects to explain for spatial heterogeneity and correlation in the data. As it has been generally observed that the standard model used in disease mapping is the Besag, York and Mollie (BYM) model which includes two random effects, with one which is spatially structured and the other random effect which is spatially unstructured and

with their respective prior distributions as conditional autoregressive (CAR) models and Gaussian distribution.

In the usual BYM (Besag, York and Mollie) model, the spatially structured random effect cannot be seen independently from the spatially unstructured random component. This has raised the problem of identifiability. These therefore made prior definitions for the hyperparameters of the two random effects thought-provoking. There are alternative model formulations that address this confounding, however, the issue on how to choose interpretable hyperpriors is still vague in most spatial modelling of the random effects. This thesis would propose a parameterisation of the BYM model that can lead to an improved parameter control as the hyperparameters can be seen independently from each other.

For the CAR or "Besag" model that is usually assumed as a prior distribution for the spatially structured random effects, some authors have also pointed out counterintuitive consequences on the prior covariance matrix of the spatial random effects (Renato *et al.,* 2009). Therefore, the need to develop a more intuitive spatially weighted ICAR prior for the spatial random component is important to enable the task of providing interpretable hyperpriors.

Chapter Three

Materials and Methodologies

3.1 Introduction

This chapter seeks to address the methods of data collection, the theories and concepts of the newly proposed methodologies that would be used to implement our objectives and in addressing our research problems. The objective of this chapter is to also develop and build the models for all the methods proposed. Alternative models to measure over-dispersion were developed, and further, a model that can measure both under-dispersion and over-dispersion was developed. Also the chapter proposes a more intuitive conditional autoregressive model (CAR), to model for spatial dependency in the BYM convolution model. A new technique from a mathematical procedure to disease mapping is also introduced in this chapter. A computationally faster estimation method, the INLA approach, to be used in the parameter estimation was also introduced. These are the contributions to the existing knowledge that this entire dissertation seek to perform.

3.2 Ethical Consideration

This study was carried out under the authorization and permission of the Ethical committee of the University of Fort Hare, Alice, Eastern Cape, South Africa and approval of the Eastern Cape Department of Health, with ethical clearance reference number QIN041SOBA01 and EC_2015RP24_398 respectively.

3.3 Data Sources

This is a retrospective secondary data source from Eastern Cape Province TB notification and survey data. All data used is an extract from the electronic tuberculosis register (ETR) records of

TB cases from the twenty-four health sub-districts of the province including the two metropolitan municipalities as shown in **Figure 3.1**. The data obtained was for the period of 2012 to 2015. The socio-economic and socio-demographic indicators and variables were obtained from publications of Eastern Cape Socio Economic Consultative Council (ECSECC 2014) reports of all the local municipalities.

3.4 Eastern Cape Province Geographical Information System

The Province of the Eastern Cape is situated on the east coast of South Africa and lies between the Western Cape and KwaZulu-Natal provinces. The Northern Cape and Free State provinces as well as Lesotho, shares borders with this Province. The Eastern Cape Province boasts of amazing natural diversity, stretching from the semi-arid Great Karoo to the forests of the Wild Coast. It also extends around the Keiskamma valley, the fertile Langkloof, and the mountainous southern Drakensberg region. The main feature of the Eastern Cape is its amazing coastline adjoining the Indian Ocean. The Province covers an area of 168 966km² and with a population of 6 562 053(Statistics South Africa, Censuses 2011). The Province is situated at 32.2968° S and 26.4194°E of the country.

The Eastern Cape is the second-largest province in South Africa by surface area and also the third-largest populated province with its capital in Bhisho. Port Elizabeth, East London, Grahamstown, Mthatha (previously Umtata), Graaf Reinet, Cradock and Port St Johns are the major towns and cities in the province. The province is divided into two metropolitan municipalities, and they are Buffalo City Metropolitan Municipality and Nelson Mandela Bay Metropolitan Municipality. It has six district municipalities and which are further subdivided into 37 local municipalities.

The Eastern Cape is regarded as one of the poorest provinces in South Africa. This is mostly as a result of the poverty found in the former homelands, where subsistence agriculture prevails.

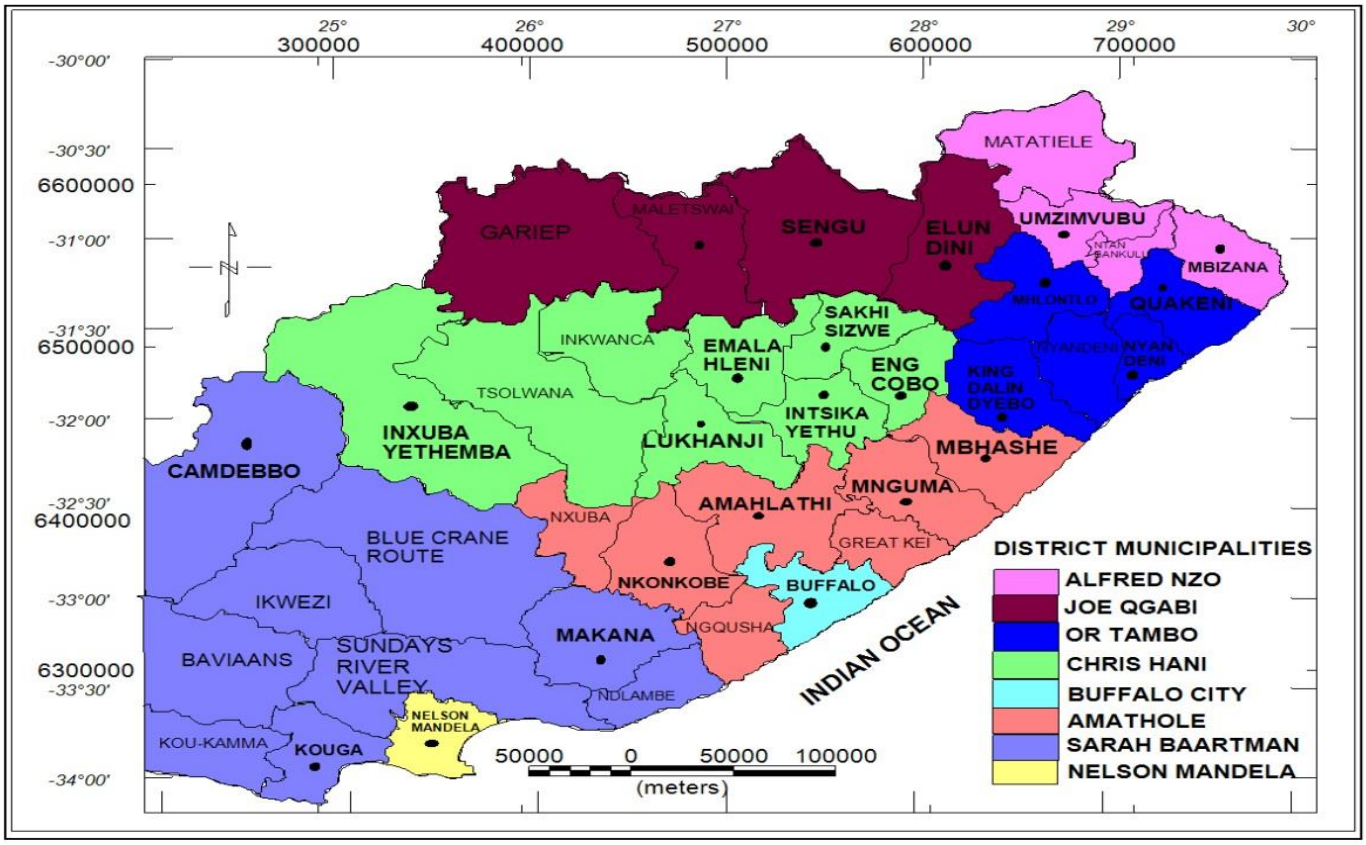


Figure 3.1: Map of Eastern Cape showing TB data registries in the 24 health sub-districts.

Source: Obaromi (2016)

3.5 Over-dispersion Model Distributions and Their Properties

3.5.1 The Poisson Model

Considering the limit of the probability mass function (*pmf*) of the Binomial distribution, as *n* tends to infinity and *p* tends to zero with $np = \lambda$, where $\lambda > 0$, then the resultant distribution has the density

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!} y = 0,1, \ldots \quad (3.1)$$

This distribution is called the Poisson distribution with parameter λ. The Poisson distribution also arises from the counting process $\{N(\tau)\}$:

3.5.1.1 Properties

- $N(0) = 0$.
- The numbers of occurrences counted in disjoint intervals are independent from each other.
- The probability distribution of the number of occurrences counted in any time interval only depends on the length of the interval.
- No counted occurrences are simultaneous.

In this process, the random variable $N(\tau)$) follows the Poisson distribution with parameter $\lambda(\tau)$, where $\lambda(\tau)$ is the rate of the event occurring in the interval time $(t, t + \tau)$ for any time t.

The Poisson density has mean and variance,

$$E(Y_i) = Var(Y_i) = \lambda_i.$$

To incorporate covariates and to ensure non-negativity, the mean or the fitted value is assumed to be multiplicative, that is, $E(Y_i|\boldsymbol{x_i}) = \lambda_i = e_i exp(\boldsymbol{x_i^T \beta})$, where e_i denote a measure of exposure, $\boldsymbol{x_i}$a $p \times 1$vector of explanatory variables, and $\boldsymbol{\beta}$ a $p \times 1$vector of regression parameters. If $\boldsymbol{\beta}$ is estimated by the maximum likelihood method, the likelihood equations are,

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta_j}} = \sum_i (y_i - \lambda_i) x_{ij} = 0, j = 1,2, \dots, p \qquad (3.2)$$

Since (3.2) is also equal to the weighted least squares, the maximum likelihood estimates, $\widehat{\boldsymbol{\beta}}$, may be solved by using the Iterative Weighted Least Squares OWLS) regression.

The index of dispersion of the Poisson distribution is thus always one.

3.5.2 The Negative Binomial Model

In single disease spatial analysis, the most commonly used model was proposed by Besag *et al.*, (1991) called the Besag, York and Mollie (BYM) model. It is used normally to model disease prevalence in areas assuming the Poisson distribution. A likely problem with such data is that over-dispersion may be evident in the data. One approach to address the problem of over-dispersion is to assume alternatively, that y_i has a negative binomial distribution (NBD), also known as the Poisson-Gamma distribution which has become the most commonly used probabilistic distribution for modelling over-dispersion in count data, with an additional

parameter called the dispersion parameter which permits the variance to exceed the mean (Bliss *et al.,* 1958; Poch *et al.,* 1996; Lord 2006; Cameron *et al.,* 1998).

The negative binomial distribution, just like the normal distribution, arises from a mathematical formula. It is normally used to describe the distribution of count data. Also like the normal distribution, it can be entirely defined by just two parameters - its mean (m) and shape parameter (k), which is commonly considered to be fixed (Lord, 2006; Bliss *et al.,* 1958), to measure over-dispersion. However, unlike the normal distribution, the negative binomial does not naturally result from the use of large samples, nor does it arise from a single causal model.

Anscombe (1950) described five (5) causal models of the negative binomial, some of which can be interpreted as due to aggregation: Inverse binomial sampling, Constant birth-death-immigration rates, Heterogeneous Poisson sampling, Compounding of Poisson and logarithmic distributions and True clumping aggregation or contagion.

3.5.2.1 Properties

Sometimes, the occurrence of an event is dependent of other events in the same sampling unit. For example, numerous subjects of interest are not distributed randomly or are not sampled randomly, and therefore, the Poisson distribution does not offer a good explanation of their pattern of dispersion. The most common pattern of spatial dispersion is usually aggregated, rather than random or regular. The same may also happen for events occurring through time. That is, an event that may 'spark off' other events and resulting in a contagious distribution. The negative binomial distribution is one of some probability distributions that can be used to describe such a pattern of dispersion.

The negative binomial is the easiest to calculate, and the most widely-applicable of over-dispersion models. Like the Poisson distribution, the negative binomial is discrete, unimodal and skewed. Statistically, its parameters are both simple and flexible. The negative binomial model can also be described as being versatile, but without carrying too deep a causative commitment.

Usually, it is used as a fairly subjective, but convenient approximation to how counts are distributed, and provided the data have a negative binomial distribution; k is used as a measure of that distribution's shape.

3.5.2.2 Negative Binomial Population Parameters

The mean, variance, skew and kurtosis of a negative binomial population can be calculated as follows:

- The **mean** frequency of failures, m, can also be calculated as 1 − k - where k is given as the mean number of successes.
- The **variance** is m(k+m)/k
- The **skew** is (1 + m/(k+m)) × √(km/(k+m))

The **kurtosis** is 3 + 6/k + k/(m(k+m))

3.5.2.3 Definition: The probability mass function (pmf) of a negative binomial regression model for independent count data observations $Y_i, i = 1,\ldots,n$ with parameters $d > 0$ and $\mu_i > 0$, represented by $Y_i \sim NB(d, \mu_i)$, is defined as

$$P\left(Y_i = y_i \mid d, \mu_i\right) = \frac{\Gamma\left(y_i + d\right)}{\Gamma\left(d\right)\Gamma(y_i + 1)} \cdot \left(\frac{d}{\mu_i + d}\right)^{d} \cdot \left(\frac{\mu_i}{\mu_i + d}\right)^{y_i}, \quad \mu > 0, d > 0, y = 0,1,\ldots \quad (3.3)$$

This distribution also arises from a Poisson distribution where the dispersion parameter d is assumed to be constant and follows a Gamma distribution with mean

$$E\left(Y_i \mid d, \mu_i\right) = \mu_i$$

and

$$\operatorname{var}\left(Y_i \mid d, \mu_i\right) = \mu_i\left(1 + \frac{\mu_i}{d}\right)$$

The parameter d with a fixed assumption over time for each region μ_i, depends on covariates through the transformation:

$$\ln(\mu_i) = \beta X_i$$

As $d \to \infty$ in the limit, the negative binomial distribution converges to the Poisson distribution with parameter μ_i (Winkelmann, 2003) and the logarithmic series distribution is obtained as $d \to 0$ (Bliss *et al.,* 1953; Anscombe, 1950). Consequently, the unobserved heterogeneity among observations can be used to interpret the over-dispersion in the negative binomial model.

3.5.2.4 Estimators of the Negative Binomial Dispersion Parameter.

From the mass function of the negative binomial distribution, it can be seen that the dispersion parameter *d,* is an essential part of the model. Estimation of *d* is thus important given a sample of counts. For example, *d* is a parameter that is critical for evolving confidence intervals and refining the forecast mean when the Empirical Bayes (EB) is used (Hauer, 1997; Wood, 2005).

In estimating the dispersion parameter (or its inverse), several estimators have been proposed. The simplest method is the Method of Moments Estimate (MME) (Anscombe, 1949).

The method of Maximum Likelihood Estimate (MLE), first proposed by Fisher (1941) and later developed by Lawless (Lawless, 1987) with the introduction of gradient elements, is also commonly used. More recently, Clark and Perry (Clark *et al.,* 1989) introduced the Maximum Quasi-Likelihood Estimate (MQLE), which is considered to be an extension of the MMEA method of multistage estimation was also presented in (Willson 1984). Among these methods, the MME and the MLE are often considered to be superior to other methods and are more widely used nowadays.

We hereby describe the two common estimators of the dispersion parameter *d*: moments, maximum likelihood estimate and just for mentioning, the recently proposed Bootstrapped maximum likelihood estimator by Zhang *et al.,* (2006). However, the full mathematical estimation of these estimators was not considered in this thesis.

3.5.2.5 Methods of Moments estimators (MME)

For a negative binomial distribution, the variance σ^2, mean μ and d have the relationship

$$\sigma^2 = \mu + \mu^2 / d \ .$$

Based on this relationship, the MME is developed and estimated by

$$\hat{d} = \frac{\bar{x}^2}{S^2 - \bar{x}}$$

Where $\bar{x}$ and S^2 are the first and second unbiased sample moments (Anscombe 1949). Note that the estimate $\hat{d}$ is reasonable only when $S^2 > \bar{x}$ because $d > 0$.

To obtain a good estimate of *d* by MME, it is very important to have a good knowledge of the variance because even a slight change of the variance may cause a large variation of *d*. This problem will be enlarged when the sample size becomes smaller.

3.5.2.6 Maximum Likelihood Estimate (MLE)

Different types of negative binomial distributions can be generated by different parametrisations (John Cook 2009). The more useful parameterization, $\alpha = 1/d$ giving

$$\mathrm{P[Y = y]} = \frac{\Gamma\left(y+\alpha^{-1}\right)}{y!\Gamma\left(\alpha^{-1}\right)}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y}\left(\frac{1}{1+\alpha\mu}\right)^{\alpha^{-1}}, \mu > 0, \alpha > 0, y = 0,1,\ldots, \tag{3.4}$$

allows for more direct identification of the dispersion parameter d. When $\alpha = 0$, note that the above distribution becomes a Poisson (μ) distribution. More normally, $\alpha < 0$ can be allowed, which suggests under-dispersion of the data and the corresponding distribution given by (3.2) becomes a Binomial distribution.

If $\alpha > 0$ (over-dispersion) is given its true value according to sampling variation under small sample sizes, the MLE of the dispersion parameter can be negative if the parameter space $\alpha \in (-1/Ymax, \infty)$ is considered (Piegorsch, 1990; Saha and Paul, 2005). The task of fitting discrete count data using the NB distribution is to find the maximum likelihood estimate (MLE) for the dispersion parameter, α.

As shown by Fisher (Fisher, 1941), the log-likelihood function from a sample of independent identically distributed (i.i.d.) variates $(x_i's)$ is proportional to

$$l(d,\mu) = \frac{1}{n}\sum_{i=i}^{n} log\left(\frac{x_i+d}{\Gamma(d)}\right) + \left[\bar{x}\log(\mu) - (\bar{x}+d)\log(1+{}^{\mu}/_{d})\right] \qquad (3.5)$$

where μ is again the mean of the NB distribution. The sample variates are integers in practice, which yields

$$\Gamma(x+d)/\Gamma(d) = (x+d-1)(x+d-2)...(d+1)k.$$

The term

$$log\left(\frac{\Gamma(x_i+d}{\Gamma(d)}\right) \qquad (3.6)$$

then can be written as

$$log\left(\frac{\Gamma(x_i+d}{\Gamma(d)}\right) = \sum_{y=1}^{x_i-1} dlog(1+v/d). \qquad (3.7)$$

without call to the gamma function (Lawless, 1987).

Thus, the log-likelihood function can finally be expressed by

$$l(d,\mu) = \frac{1}{n}\sum_{i=1}^{n}\sum_{v=1}^{x_i-1}\mathrm{d}\log(1+v/d) + \left[\bar{x}\log(\mu) - (\bar{x}+d)\log(1+u/d)\right] \qquad (3.8)$$

with gradient elements

$$\nabla_u l = \frac{\bar{x}}{u} - \frac{1+\bar{x}/d}{1+u/d} \text{ and } \nabla_d l = \frac{1}{n}\sum_{i=1}^{n}\sum_{v=1}^{x_i-1}\frac{v}{1+v/d} + d^2 \log\left(1+\frac{u}{d}\right) - \frac{u(x_i+d)}{1+\frac{u}{d}}. \quad (3.9)$$

From the gradient elements, $\nabla_u l = 0$ yields $\hat{\mu} = \bar{x}$. Then the MLE of d can be obtained via a nonlinear root-finder by setting $\nabla_d l = 0$ and given $\mu = \hat{\mu}$.

Although a vast literature exist on how to estimate the dispersion parameter α (Willson *et al.*, 1984; Clark and Perry, 1989; Piegorsch, 1990; Saha and Paul, 2005) and several estimation approaches are given. All current works showed that finding the MLE for α is a challenging one, since the roots of the score function can be no root, one root or more than one root. The index of dispersion is given by 1/d > 1 and the negative binomial distribution is thus always overdispersed.

3.5.3 The Generalized Poisson distribution

The Generalized Poisson Distribution (GPD) was introduced in Consul and Jain (1973) and studied extensively by Consul (1989). The generalized Poisson (GP) distribution when defined using the maximum likelihood estimation methods for its parameters, has the probability density function (Wang *et al.*, 1997),

$$\Pr((Y_i = y_i) = \left(\frac{\mu_i}{1+a\mu_i}\right)^{y_i} \frac{(1+ay_i)^{y_i-1}}{y_i!} exp\left(-\frac{\mu_i(1+ay_i)}{1+a\mu_i}\right), \quad y_i = 0,1,\ldots \quad (3.10)$$

with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = \mu_i(1+a\mu_i)^2$.

The Generalized Poisson (GP) is a natural extension of the Poisson distribution. If a equals zero, the Generalized Poisson reduces to the Poisson distribution, resulting into $E(Y_i) = Var(Y_i)$.

If $a > 0$, the variance is larger than the mean, $Var(Y_i) > E(Y_i)$, and the distribution represents count data with over-dispersion. If $a < 0$, the variance is smaller than the mean, $Var(Y_i) < E(Y_i)$, so that now the distribution represents count data with under-dispersion. If by assumption, the mean or the fitted value is multiplicative, that is, $E(Y_i|\boldsymbol{x_i}) = \mu_i = e_i exp(\boldsymbol{x_i^T \beta})$, the likelihood for Generalized Poisson regression model may be written as,

$$l(\boldsymbol{\beta}, a) = \sum_i y_i log\left(\frac{\mu_i}{1+a\mu_i}\right) + (y_i - 1)log(1 + a\mu_i) - \frac{\mu_i(1+ay_i)}{1+a\mu_i} - log(y_i!) \qquad (3.11)$$

Therefore, the maximum likelihood estimates, $(\widehat{\boldsymbol{\beta}}, \hat{a})$, may be obtained by maximizing $l(\boldsymbol{\beta}, a)$with respect to $\boldsymbol{\beta}$and a. The associated equations are,

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta_j}} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{(1+a\mu_i)^2} = 0, \qquad j = 1,2, \dots, (3.12)$$

and,

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{a}} \sum_i - \frac{y_i\mu_i}{1+a\mu_i} + \frac{y_i(y_i-1)}{1+ay_i} - \frac{\mu_i(y_i-\mu_i)}{(1+a\mu_i)^2} = 0. \qquad (3.13)$$

The main advantage of using the Generalized Poisson distribution is that it can be fitted for both over-dispersion, $Var(Y_i) > E(Y_i)$ as well as under-dispersion, $Var(Y_i) < E(Y_i)$, which serves as an advantage over the negative binomial distribution.

3.6 Modelling the Prior for the Spatial Dependency Structure.

3.6.1 The CAR "Besag" model for the structured spatial component.

As defined in the literature, the CAR prior model is an intrinsic Gaussian Markov random field (GMRF) model, here referred to as the Besag ICAR model and it is one of the most popular approaches to model spatial correlation.

The Besag (CAR) model for random vector $x = (x_1, \dots, x_n)$ is defined as

$$x_i | x_j, i \neq j, \tau \sim N\left(\frac{1}{n_i}\sum_{i\sim j} x_j, \frac{1}{n_i \tau}\right) \qquad (3.14)$$

where n_i is the number of neighbours of node i, $i \sim j$indicates that the two nodes i and j are neighbours. The mean of x_j equals the mean of the effects over all neighbours, and the precision n is proportional to the number of neighbours.

Hyperparameters

The parameter τ is the precision parameter and it is represented as

$$\theta_1 = log\ \tau$$

and the prior is defined on θ_1.

3.6.2 A novel Besag2 ICAR Model for the Structured Spatial Effects.

A modification of the Dean model in (2.42) was proposed by Simpson *et al.*, (2015) and which addresses both the identifiability and scaling issue of the BYM model. The model uses a scaled structured component μ_{i*} where τ is the precision matrix of the Besag ICAR model.

The "Besag2" is one of the models in the latent Gaussian field. The Besag2 model is an extension to the Besag (ICAR) model above in (3.14).

Paramerisation of the Besag2 model

Let the random vector $z = (x_1, \dots, x_n)$ be the "Besag" model (ICAR), then the "Besag2" is the following extensions

$$x = (az, \frac{z}{a}) \tag{3.15}$$

where $a > 0$ is an additional hyperparameter and $\dim(x) = 2n$, and z is the same (a tiny additive noise) random vector.

Hyperparameters

This model has two hyperparameters $\theta = (\theta_1, \theta_2)$. The precision parameter τ is represented as

$$\theta_1 = log\tau \tag{3.16}$$

And the prior is defined on θ_1.

The weight parameter a is signified as

$$\theta_2 = \log a \tag{3.17}$$

And the prior is defined on θ_2. The precision defines how equal the two copies of z is.

This new prior is a member of the class of a latent Gaussian (LGMs) markov random field models and would be compared with the Besag ICAR, and the BYM models for flexibility and robustness.

3.7 Proposed Model Building and the Covariates

3.7.1 Statistical Analysis

Our working model is the usual BYM model (Besag *et al.*, 1991) which is a type of generalized linear mixed model (GLMM) with both fixed and random effects;

$$logy_i = \alpha + \sum_{i=1}^{k} \beta_i x_{ik} + \boldsymbol{b} \qquad (3.18)$$

where $\boldsymbol{b}$ is the random effects which are further broken into spatially structured, u_i, and spatially unstructured, v_i, random effects.

Modifications and adjustments would be made on the BYM model and applied to observed TB counts in region i to estimate the relationship between TB relative risks and some explicative variables of socio-economic vulnerability and demographic factors. Also on this working model, all other model specifications and comparisons would be carried out by adjusting for spatial and non-spatial components. We adopted seven (7) different full hierarchical Bayesian models to estimate the relative risk of the occurrence of TB through the integrated nested Laplace approximation (INLA) method.

3.8 Definition of the Covariates in the Regression Model

For our model, six covariates (socio-economic and demographic factors) were taken as explanatory variables for the relative risk of the disease. Let λ_i be the number of new TB cases in area *i*, x_1 = Gini coefficient (a measurement of how income or poverty is equally distributed), x_2 = poverty rate (% number of individuals living below the poverty line. Though, there is no official poverty line defined for South Africa), x_3 = unemployment rate (%), x_4 = No schooling (%; person aged 20+ years), x_5 = average household size and x_6 = Population density of the regions/municipalities. The Gini coefficient, poverty rate and unemployment are considered in this study as distal factors, while no schooling, average household size and population density are taken as proximal factors. Two spatial random effects of u_i and v_i respectively as unstructured random effect to measure for spatial heterogeneity and a structured random effect to measure for spatial dependency among the regions.

3.8.1 Fitting the Spatial Multilevel Models

For the analysis of the estimated risk factors to determine the effects of these covariates, our proposed regression model above will give rise to seven (7) multilevel or hierarchical Bayesian models in the following order:

Model 1:; $log\lambda_i = \beta_0 + \sum_{i=1}^{3} Xi\beta_i$:$O_i \sim Poisson\ (\lambda_i)$ [analyse with only the distal factors without the spatial random effects].

Model 2: $log\lambda_i = \beta_0 + \sum_{i=4}^{6} Xi\beta_i$: $O_i \sim Poisson\ (\lambda_i)$ [analyse with only the proximal factors without the spatial random effects].

Model 3: $log\lambda_i = \beta_0 + \sum_{i=1}^{6} Xi\beta_i$: $O_i \sim Poisson\ (\lambda_i)$ [analyse with both the distal and proximal factors without the spatial random effects].

Model 4: $log\lambda_i = \beta_0 + \sum_{i=1}^{3} Xi\beta_i + v_i + u_i$; O_i~*Poisson*(λ_i), μ_i ~*ICAR*, v_i~N$(0, \sigma_v^2)$ [analyse with only distal factors and spatial random effects]

Model 5:; **:**$log\lambda_i = \beta_0 + \sum_{i=4}^{6} Xi\beta_i + v_i + u_i$; O_i ~*Poisson* (λ_i), μ_i ~*ICAR*, v_i~N$(0, \sigma_v^2)$ [analyse with only proximal factors and spatial random effects] .

Model 6: $log\lambda_i = \beta_0 + \sum_{i=1}^{6} Xi\beta_i + v_i + u_i$; $O_i \sim Poisson\ (\lambda_i)$, μ_i ~*ICAR*, $v_i \sim$ N$(0, \sigma_v^2)$ [analyse with both the distal and proximal factors with the spatial random effects]

Model 7: $log\lambda_i = \beta_0 + v_i + u_i$:$O_i \sim Poisson\ (\lambda_i)$, μ_i ~*ICAR*, $v_i \sim$ N$(0, \sigma_v^2)$[the model with only the spatial random effects].

3.9 MODEL ASSUMPTIONS AND ESTIMATION

Model estimations were carried out following Bayesian techniques and appropriate priors were assigned to all the functions and terms. A non-informative prior knowledge was considered with a flat distribution for the intercept β_0. For the β_i, we assumed non-informative Gaussian priors with zero mean and precision equal to 1×10^{-5}. The spatially structured random effects $u_i = (u_i, \ldots, u_n)$ – accounts for the spatial dependence, with prior distribution taken as a conditional intrinsic Gaussian autoregressive model, ICAR. Also in this context, the non-spatial or unstructured spatial random effects was assumed to have an independent Gaussian distribution of zero mean and variance σ_v^2 following an inverse gamma distribution as $1/\sigma_v \sim dgamma(0.5, 5 \times 10^{-4})$.

The entire above model fitting and analyses would be done by first assuming a Poisson distribution for λ_i as seen above and thereafter, the proposed negative binomial and generalized Poisson distributions respectively, for the observed counts.

3.10 Bayesian Modelling Approach

Bayesian analysis rests upon computing the posterior probability distribution for model parameters. The posterior probability distribution is the conditional probability distribution of the unknown parameters, given the observed data and weighted by the prior information. Bayesian modelling depends on the ability to compute posterior distributions in order to provide estimates for all the corresponding model parameters. Majority of these posterior distributions are straightforward to calculate. Distributions with a conjugate prior typically have a posterior distribution which follows a standard distributional form.

Bayesian Inference

The prior distribution is the distribution of the parameter(s) before any data is observed, that is,

$$p(\theta|\alpha).$$

The prior distribution might not be easily determined. In this case, we can use the Jeffreys prior to obtain the posterior distribution before updating them with newer observations.

The sampling distribution is the distribution of the observed data conditional on its parameters, i.e.

$$p(X|\theta)\ .$$

This is also termed the likelihood, especially when viewed as a function of the parameter(s), sometimes written,

$$L(\theta|X) = p(X|\theta).$$

The marginal likelihood (sometimes also termed the evidence) is the distribution of the observed data marginalized over the parameter(s),

$$p(X|\alpha) = \int_{\theta} p(X|\theta)\, p(\theta|\alpha) d\theta$$

The posterior distribution is the distribution of the parameter(s) after taking into account the observed data. This is determined by Bayes' rule, which forms the heart of Bayesian inference

$$p(\theta|X,\alpha) = \frac{p\big(X\big|\theta\big)\, p(\theta|\alpha)\big)}{p(X|\alpha)} \propto\ p(X|\theta)p(\theta|\alpha)$$

In many cases, however, the computation required is more complex and a more advanced method is essential to calculate the posterior distribution. These advanced approaches usually make use of some form of numerical simulation, generally by drawing a sample of parameter values from an approximation of the posterior distribution $f(\theta|Y)$ to allow estimation of the distributions of the model parameters.

3.11 Estimation by the Integrated Nested Laplace Approximation (INLA) method.

The posterior marginals are not always presented in closed form as a result of the non-Gaussian response variables. For such models, Markov chain Monte Carlo methods can be applied, but they are not without some problems, both in terms of convergence and in computational time. In some practical uses, the extent of these problems is such that Markov chain Monte Carlo is simply not an appropriate tool for monotonous analysis.

It is shown in that by using an integrated nested Laplace approximation (INLA) and its simplified version; we can directly compute very precise approximations to the posterior marginals. The key advantage of these approximations is simply computational: where MCMC algorithms need hours and days to run, INLA provide more precise estimates in seconds and in minutes.

Another benefit with INLA approach is its generality, which makes it possible to perform Bayesian analysis in a programmed, streamlined way, and to compute model comparison criteria and various predictive measures so that models can be compared and the model under study can be tested. This method is also used where the model has a hidden Gaussian Markov Random field, with the parameters of interest being latent variables which are not observed directly, but are instead inferred from other observed variables.

Considering the following hierarchical model,

$$Y_i = Poisson\,(\mu_i) \quad i = 1, \ldots, n$$

$$\log(\mu_i) = x_i^T \beta + \phi_i \tag{3.19}$$

Given that, $\phi = (\phi_1, ..., \phi_n)$ comprise a set of random effects, which can be considered as a group of latent variables. Let ω be the set of hyperparameters relating to ϕ, then the marginal posterior for each variable ϕ_i is as follows:

$$\pi(\phi_i \mid Y) = \int_{\omega} \int_{\phi^{-1}} \pi(\phi, \omega \mid Y) d\phi_{-i} d\omega \tag{3.20}$$

where, ϕ^{-1} is the vector ϕ, with element ϕ_i removed.

This can be modified as

$$\pi(\phi_i \mid Y) = \int_{\omega} \pi(\phi_i \mid \omega, Y) \pi(\omega \mid Y) d\omega \tag{3.21}$$

INLA involves the construction of a nested approximation of (3.15), which requires approximations of $\pi(\omega \mid Y)$ and $\pi(\phi_i \mid \omega, Y)$.

Here $\pi(\omega \mid Y)$ can be approximated using the following Laplace approximation

$$\tilde{\pi}(\omega|Y) \propto \frac{\pi(\phi, \omega, Y)}{\tilde{\pi}_c(\phi_i|\phi, \omega, Y)} |\phi = \phi^*(\omega) \tag{3.22}$$

where $\tilde{\pi}_c(\phi_i|\phi, \omega, Y)$ is termed the Gaussian approximation to the full conditional distribution of ϕ and $\phi^*(\omega)$ is the mode of the full conditional distribution of ϕ for a given value of ω.

The authors in Rue *et al.*, (2009) propose using a Laplace estimate of $\pi(\phi_i \mid \omega, Y)$ which takes the following form:

$$\tilde{\pi}_{LA}(\phi i|\omega, Y) \propto \frac{\pi(\phi,\omega,Y)}{\tilde{\pi}_G(\phi_{-i}|\phi_i,\omega,Y)}|\phi_{-i} = \phi^*_{-i}(\phi_i,\omega) \qquad (3.23)$$

where $\tilde{\pi}_G(\phi_{-i}|\phi_i,\omega,Y)$ is the Gaussian approximation to $\phi_{-i}|\phi_i,\omega,Y$ and $\phi^*_{-i} = (\phi_i,\omega)$ as its modal value for a given ω.

Even though in most cases, MCMC and INLA inferences give the same and related results, it should be noted that there are vital differences in the way that posterior distributions are estimated. MCMC is capable of sampling directly from a joint posterior distribution, while INLA on the other hand, uses a closed form expression to estimate the marginal posterior distributions. In MCMC inference, the joint posterior distribution can be estimated directly, and will take the form of a straight line which passes through the points. In contrast, INLA estimates the marginal distributions individually, and both of these will be alike to the prior. For this thesis, we adopted the latter.

3.12 Model Comparison

The comparison of numerous contending Bayesian models is usually a challenging task and needs special attention (see Kass and Raftery, 1995). Since the models used include sets of random effects, the Deviance Information Criterion (DIC) shall be used for comparison (Spiegelhalter *et al.*, 1998).

It is defined as DIC = $\bar{D} + pd$, where $\bar{D} = E[-2\log(\hat{L})]$ is given as the mean posterior deviance and *pd* represents the actual number of parameters. When two or more models are compared, the model with the least DIC value would be adopted.

Similar to the BIC, this approach penalises models which have superfluous parameters, and favours approaches which provide a sensible data fit while minimising the amount of parameters.

3.13 Geospatial Disease Mapping Using the Biharmonic Splines

It is important to examine how a disease prevalence rates are distributed in space and how they relate to each other within a defined distance and direction. It is usually anticipated that regions or locations that are close neighbours tend to have similar rates than regions or locations that are far apart, given that the socio-economic and demographic behaviours may exceed geographical boundaries.

For the geographic and graphic representation of the disease prevalence, linear and biharmonic curve fitting methods in MATLAB 7.10.0 would be used to identify and localize areas of TB clusters. The biharmonic spline interpolation methods, which is based on Green's function and proposed by Sandwell (1987), has become the conventional technique for its high precision, simplicity and flexibility. This method would be employed in describing the spatial distribution of the disease, other than the traditional GIS mapping software. Spline Smoothing is best effective and operational for eliminating angular contours or surfaces by filling in a sparse grid.

Surface and contour plots would be produced for TB incidence at the provincial level for the period 2012-2015. Smooth functions of the prevalence rates by the coordinates-longitudes and latitudes, using biharmonic (v4) method in the software would be fitted.

The provincial TB data for 2012, 2013, 2014 and 2015 were used as the separate outcome variables, then the curve fitting plots to describe the spatial patterns and distribution of tuberculosis cases in the province. In this thesis, all computational and mathematical expressions of this method were disregarded.

Chapter Four

Results and Interpretations

This chapter focuses on the results obtained from the methodologies developed and adopted and also followed respectively by the interpretations. The results are outlined in two sections: Statistical and spatial/geospatial assessment. The statistical analysis is to show the descriptive statistics and the posterior estimates from the Bayesian approach using the INLA method. This chapter focus mainly on parameter estimations and comparisons of all the proposed models. Also the spatial/geospatial assessment would display the spatial patterns and distribution of the disease incidence and prevalence in the twenty four health sub districts in the province.

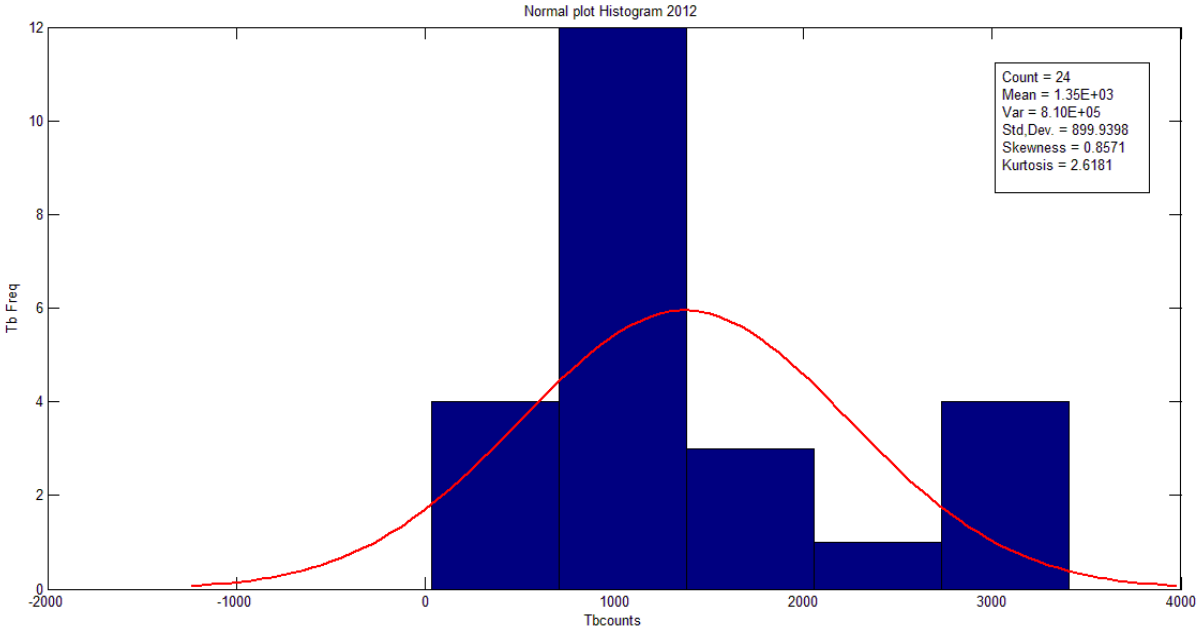
4.1 Statistical Analyses

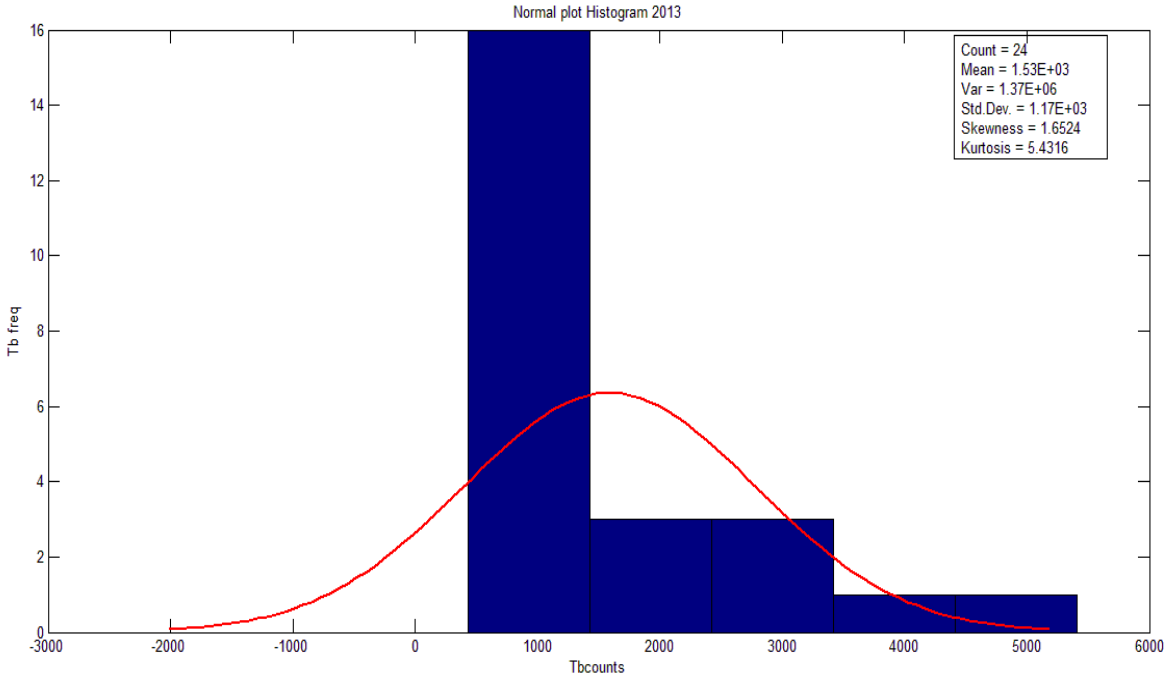
4.1.1 Descriptive Analyses

Table 4.1 and figure 4.1 shown below displayed the histograms and descriptive statistics of TB cases respectively. It shows the mean, variance, standard deviation, skewness and kurtosis of the data. This was carried out to test for normality in the data, and it could be seen that all the distributions of the data are skewed to the right and also showed a high degree of dispersion in the mean-standard deviation ratio.

Table 4.1: Table of Descriptive Statistics for TB Counts from 2012-2015.

Descriptive Statistics				
N = 24	**2012**	**2013**	**2014**	**2015**
MEAN	1.35E+03	1.53E+03	1.51E+03	1.54E+03
VAR	8.10E+05	1.37E+06	1.39E+06	1.29E+06
STD. DEV	899.9398	1.17E+03	1.18E+03	1.13E+03
SKEWNESS	0.8571	1.6524	1.7746	1.6234
KURTOSIS	2.6181	5.4316	5.7433	4.8264


Normal plot Histogram 2012
Count = 24
Mean = 1.35E+03
Var = 8.10E+05
Std.Dev. = 899.9398
Skewness = 0.8571
Kurtosis = 2.6181
Tb Freq
Tbcounts
0
2
4
6
8
10
12
-2000
-1000
0
1000
2000
3000
4000


Normal plot Histogram 2013
Count = 24
Mean = 1.53E+03
Var = 1.37E+06
Std.Dev. = 1.17E+03
Skewness = 1.6524
Kurtosis = 5.4316
Tb freq
Tbcounts
0
2
4
6
8
10
12
14
16
-3000
-2000
-1000
0
1000
2000
3000
4000
5000
6000

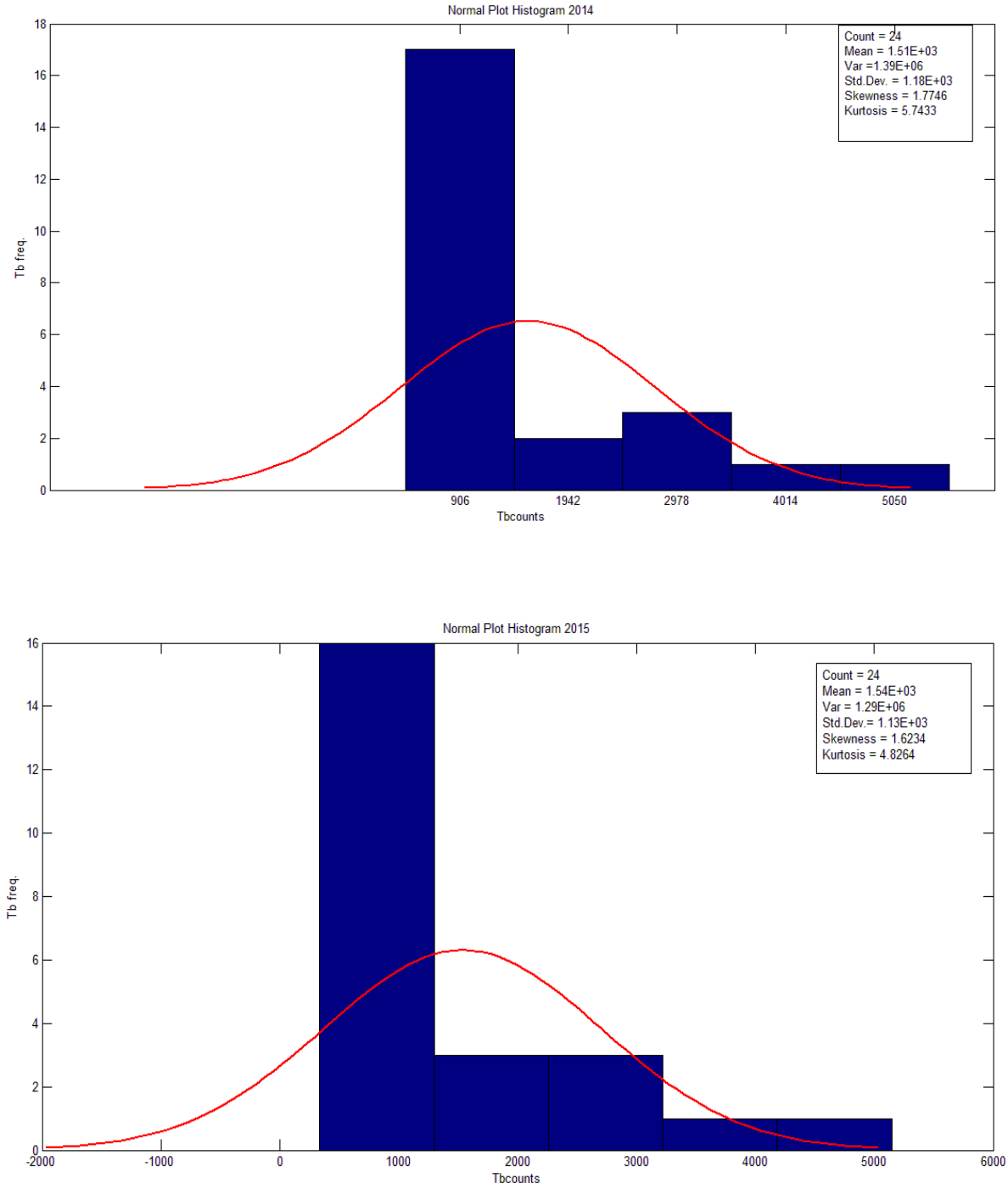
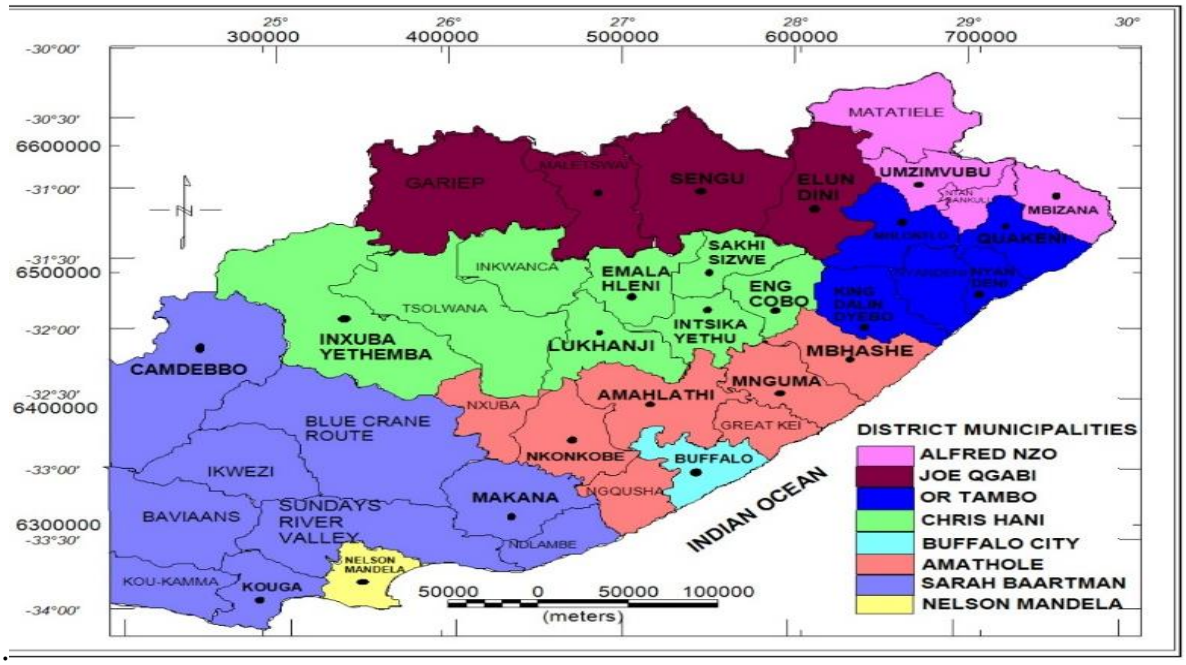
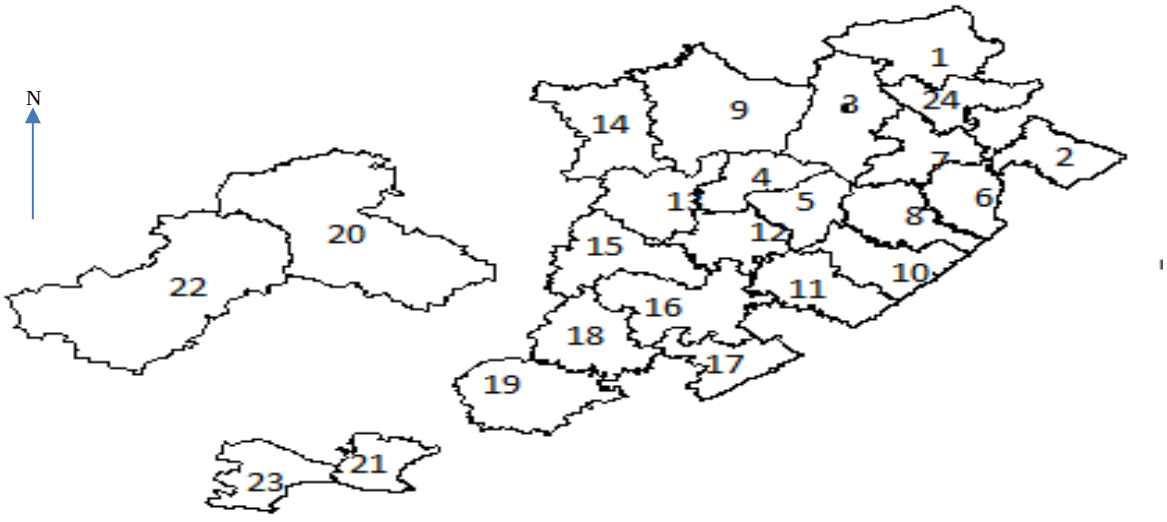


Figure 4.1: Normal histogram Plots of TB cases for 2012-2015.

4.2 Bayesian Spatial Analyses



(a).



(b)

Figure 4.2: Map of (a). Eatsern Cape Province showing the 37 local municipalities and the 2 metros and (b). extracted map showing the 24 health sub-districts for TB dataset.

Table 4.2: List of locations of TB registries in the Eastern Cape province, South Africa.

S/N	LOCATION	13	Emalahleni
1	Matatiele	14	Maletswai
2	Mbizana	15	Lukhanji
3	Elundini	16	Amahlathi
4	Sakhisizwe	17	Buffalo City
5	Engcobo	18	Nkonkobe
6	Nyandeni	19	Makana
7	Mhlontlo	20	Inxuba Yethemba
8	King Dalindyebo	21	Nelson Mandela
9	Sengu	22	Camdebbo
10	Mbhashe	23	Kouga
11	Mnguma	24	Umzimvubu
12	Intsika Yethu		

Table 4.3 below showed the estimates from the combination of Bayesian approach and GLMM to assess the spatial heterogeneity in the TB relative risk for 2014 dataset, by investigating its relationship with some socio-economic and demographic variables in the Eastern Cape Province of South Africa. For 2014 data, Eastern Cape had 37,365 notified TB cases from all the twenty-four (24) health sub-districts and with about 91.1% bacteriological coverage (ratio of number of PTB patients diagnosed by bacteriological tests to the total PTB patients reported, excluding children 0-4 years).

In this study, seven (7) separate multilevel models including/excluding the covariates/spatial random effects were developed and treated as non-independent Poisson random variables with means $\lambda_i = (\lambda_1, \dots, \lambda_n)$, to investigate whether the covariates influenced part or all of the spatial correlation in the TB relativity. In **Table 4.3**, the measures of association and their respective means and standard errors are presented for all covariates in all models. In the distal models, the most significant explanatory variable is poverty. Gini coefficient, unemployment and no schooling, which are one of the most widely used socio-economic indicators in South Africa, were not significant in any of the models and with very low standard errors. They are possibly due to either indirect effect of the trio in TB occurrence, through their effects on income (Singh-Manoux *et al.,* 2002). For the proximal models in model 2 and the rest of the models, average household size was significant, indicating that an increase of one more person in the house increases the risk of TB. Population density also play a little significant role in the disease incidence in the province (models 2, 3 5 and 6), and which indirectly influences the average household number of persons living together. In the models without the spatial random effects (models 1, 2 and 3), the effects of poverty and average household size were vital. The largest significant factor, however, is the average household size per house.

Table 4.3: Posterior Estimated (means and std. deviations) of Risk Factors Associated with TB by GLMM Model for Eastern Cape Province.

Covariates	Model 1 Distal model	Model 2 Proximal model	Model 3 Proximal and distal	Model 4 Distal with random effects	Model 5 Proximal and random effects	Model 6 Proximal, Distal and random effects	Model 7 Spatial random effects
Intercept	8.31 (0.035)	5.99 (0.062)	7.19 (0.072)	8.08 (0.760)	6.68 (0.422)	7.38 (1.071)	7.13 (0.132)
Gini coeff.	-0.33 (0.042)		-0.81 (0.045)	-0.60 (0.852)	-	-0.65 (0.647)	-
Poverty	0.10 (0.032)		0.08 (0.038)	0.33 (0.840)	-	0.06 (0.674)	-
Unemploym.	-0.03 (0.001)		-0.02 (0.001)	-0.02 (0.012)	-	-0.02 (0.009)	-
Pop. Dens.	-	0.002 (0.000)	0.002 (0.000)		0.005(0.000)	0.003 (0.001)	-
No school		-496.34(16.00)	-388.89 (17.128)		-5.17(31.302)	-5.40 (31.512)	-
Ave hhold size	-	0.44 (0.020)	0.35 (0.021)		0.02 (0.116)	0.10 (0.266)	-
ui-struc std	-	-	-	18369.33	1.21	1.55	1.841e+04
vi-unstruc std	-	-	-	0.84	18373.35	18342.02	7.319e-01
pD	**4.51**	**4.20**	**7.10**	**23.95**	**23.80**	**23.92**	**23.95**
DIC	**13853.45**	**7123.56**	**6206.43**	**263.17**	**262.99**	**263.14**	**263.16**
Log L	**-7186.85**	**-3848.68**	**-3336.43**	**-219.18**	**-211.30**	**-227.39**	**-206.25**

Although, it is an indicator of poverty and population density, it is also an important covariate in the household transmission of TB. The addition of spatial random effects to the full model comprising the distal and proximal models decreased the significance of the effects of the previous covariates on TB incidence. By adding the CAR random effects, we also found that they can bring about large changes in the posterior mean and variance of fixed effects compared to the non-spatial regression model. On the basis of DIC, model 5 performed better than all the other models, in which only the proximal factors and random effects were considered. In this best fit model, average household size and population density had a positive association with the relative risk of TB prevalence in the province. The impression is that models with smaller DIC should be chosen to models with larger DIC. Models are penalized both by the value of $\overline{D}$, which favours a good fit, but also (in common with AIC and BIC) by the effective number of parameters, pD. Since $\overline{D}$ will decrease as the number of parameters in a model increases, the pD term recompenses for this effect by favouring models with a smaller number of parameters.

When the GLMM is adjusted for the proximal and distal spatial random effects models, it is seen here to affect largely only the effect of poverty and which decreased the effect of almost all the other covariates. This outcome can be understood not only as a result of the colinearity between the average household size and poverty, but because of the fact that poverty is spatially clustered.

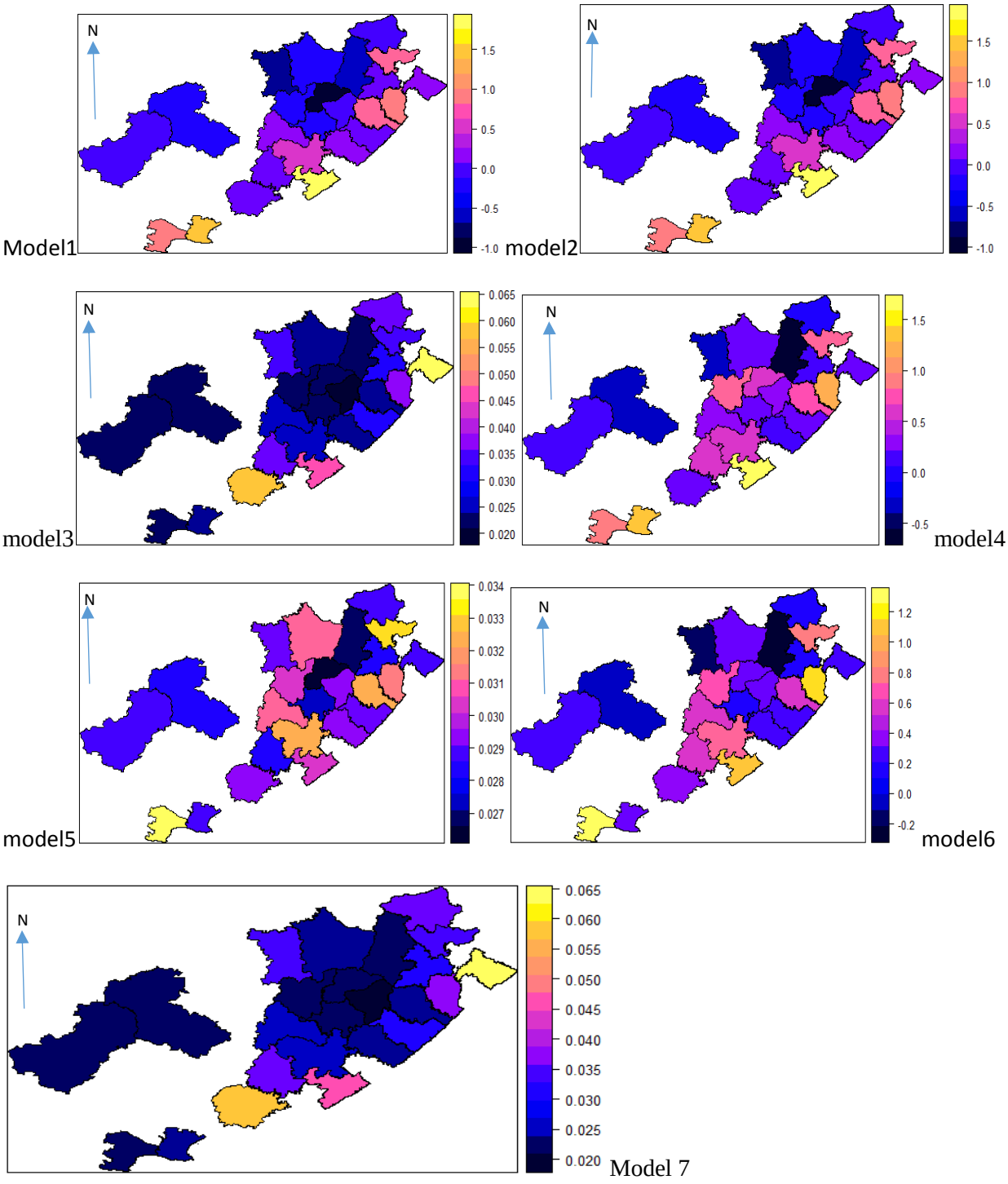


Figure 4.3: Maps of Estimated Relative Risks from the Seven (7) Models Using the Poisson Model.

4.2.1 Modelling Over-Dispersion Using Negative Binomial (NB) and the Generalized Poisson (GP) Distributions.

We hereby compare the mixture models (NB and GP) with Poisson model from the best fit model (model 5) to ascertain how they capture over-dispersion and also see how they affect the entire estimates from the Poisson regression model.

From model 5 which gave the best fit model on the basis of DIC, we have

$log\lambda_i = \beta_0 + \sum_{i=4}^{6} Xi\beta_i + v_i + u_i; O_i \sim Poisson\ (\lambda_i),\ \mu_i \sim ICAR,\ v_i \sim N(0, \sigma_v^2)$ [analyse with only proximal factors and spatial random effects].

In order to compare the presented models in **Table 4.4** and **Table 4.5**, the DIC, the posterior means of the over-dispersion parameters and the effective number of parameters are given for each model.

For the models with spatial random effects, the lowest value of the DIC is obtained for the Poisson model, while the DIC for the NP and GP model takes higher values. Hence, according to the DIC, the Poisson model is considered best among the spatial random effect models, while the NB and GP models clearly performed lesser. But considering the variances of the unstructured spatial effects for the three models, GP has the lowest variance and therefore can be said to be the best fit model on this basis. As we know that the lesser the variance, the better the precision. The effective number of parameters *pD* is close to the true number of parameters which is twenty four for the Poisson regression model, four for the NB and two for the GP regression model.

Also, with the inclusion of spatial random effects, the posterior mean of the deviance and the number of effective parameters in the NP model hardly changed, but changes were observed in the GP models.

For the Poisson model, a significant drop in the DIC is observed when spatial effects are taken into account and a slight decrease in the NB model. This demonstrates that there are some extra variability in the data which is not adequately described by the covariates only in these models. For the non-spatial models, the lowest values of the DIC are obtained for the NP and GP models, while the DIC for the Poisson takes the highest value. Hence, according to the DIC, the NP and GP models are considered best among the models without spatial random effects, while the Poisson model clearly performed worse. Comparing the NB and GP models without spatial random effects, the NB model slightly performed better than the GB model on the basis of DIC, but GP would be preferred over NB on the basis of standard error estimates as GP model estimates tend to have lesser variance values. Also the effective number of parameters *pD* is close to the true number of parameters in the models.

Since the Poisson model does not allow for over-dispersion and the heterogeneity is not of a zero inflated nature, therefore, for these two models (NB and GP), the unexplained variability is covered by the spatial effects. According to the DIC, the spatial random effect Poisson model gave the best fit and it is to be preferred to the non-spatial NB and GP models. It should be noted also, that the DIC must be used with care in this case, since firmly speaking, the DIC is defined for distributions of the exponential family only.

Table 4.4. Results to Compare Poisson, NB and GP Models in Handling Over-Dispersion with the inclusion of Spatial Random Effects.

Covariates	Poisson	NB	GP
Intercept	6.68 (0.422)	6.96 (0.990)	3.92 (0.564)
Pop. Dens.	0.005(0.000)	0.004 (0.001)	0.003 (0.001)
No school	-5.17(31.302)	-5.93 (31.538)	1.09 (31.453)
Ave hhold size	0.02 (0.116)	-0.02 (0.267)	0.79 (0.138)
ui-struc std	1.21	1.840e+04	18372.43
vi-unstruc std	18373.35	1.841e+04	18095.99
Mean of over-dis.	-	3.412	32.38
pD	**23.80**	**3.761**	**1.797**
DIC	**262.99**	**382.87**	**396.14**
Log L	**-211.30**	**-208.21**	**-301.36**

Table 4.5. Results to Compare Poisson, NB and GP Models in Handling Over-Dispersion without the inclusion of Spatial Random Effects.

Covariates	Poisson	NB	GP
Intercept	5.99 (0.062)	6.96 (1.005)	6.41 (0.796)
Pop. Dens.	0.002 (0.000)	0.004 (0.001)	0.003 (0.000)
No school	-496.34(16.00)	-5.80 (31.544)	-8.44 (31.389)
Ave hhold size	0.44 (0.020)	-0.02 (0.271)	0.15 (0.212)
Mean of over-dis.	-	3.412	14.14
pD	**4.20**	**4.048**	**3.534**
DIC	**7123.56**	**383.47**	**384.40**
Log L	**-3848.68**	**-206.62**	**-219.28**

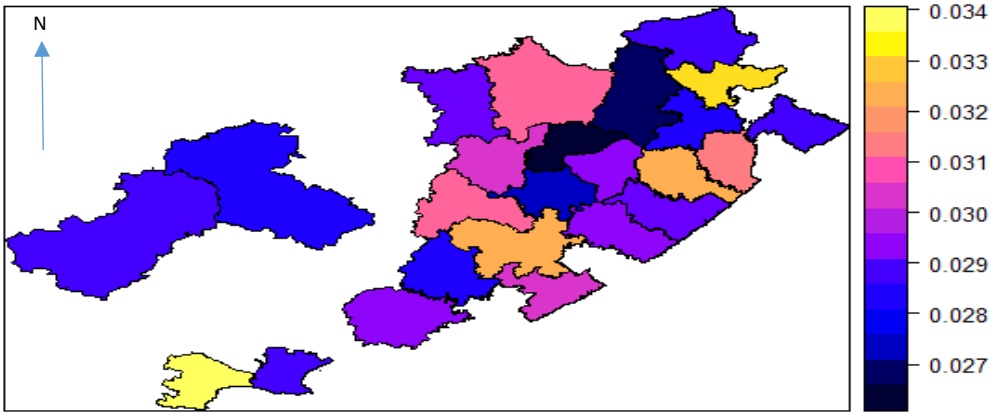
By contrast of both methods for modelling over-dispersion (spatial and non-spatial models), it is seen that all the model distributions (Poisson, NB and GP) gave better fits when the spatial random effects were included. . Also by adding the spatial random effects can cause some changes in the posterior means and variances of fixed effects compared to the non-spatial random effect regression model.

Interestingly, the results showed some differences and effects of the standard errors on the parameter estimates, based on these model distributions. Based on the comparison between Poisson, Negative Binomial and Generalized Poisson additive regression models, the Poisson, Negative Binomial and the Generalized Poisson produce almost similar estimates for the regression parameters in both models (spatial and non-spatial), but the standard errors for the Negative Binomial and the Generalized Poisson are larger than the Poisson.

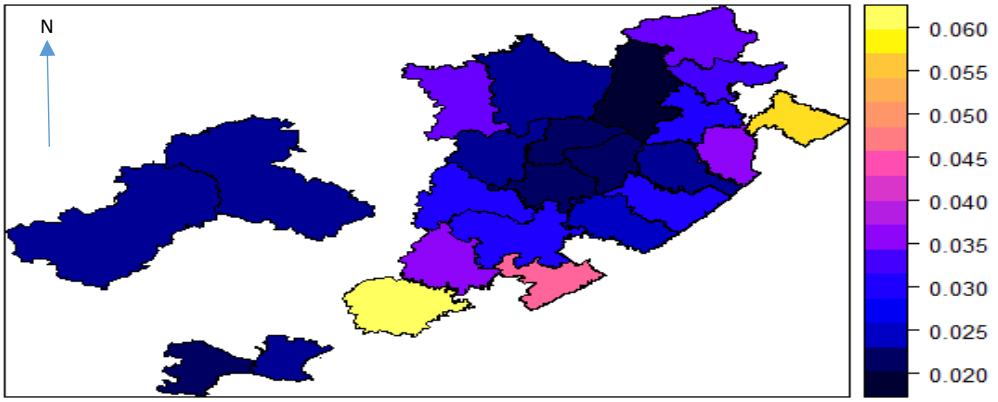
Hence, the Poisson can be said to overstate the significance of the regression parameters in the presence of over-dispersion.

The spatial maps in **Figure 4.4** below showed varying patterns of TB distribution in the province using the three models (Poisson, NB and GP). The NB and GP spatial maps displayed a more systematic pattern of TB incidence in the twenty four sub-districts than the one displayed by the Poisson map, which showed a somewhat dispersed disease pattern, hence NB and GP modelling of the TB counts gave a more detailed and clustered disease patterns.

(1).



(2).



(3).

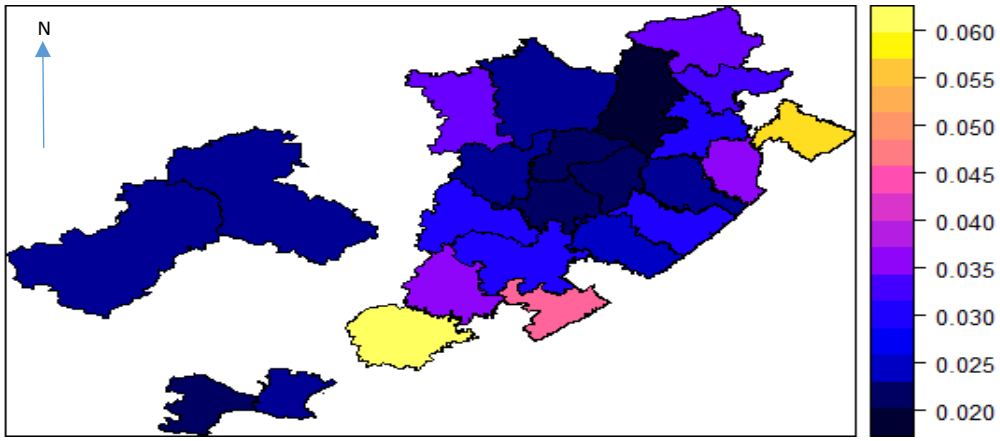


Figure 4.4. Posterior Spatial Maps At 97.5% C.I for Comparing (1) Poisson Model (2). NB Model (3). GP Model.

4.2.2 Modelling the Structured Random Effects Using the Extended Weighted "Latent" ICAR Prior.

For this analysis, the standard Besag, York and Mollie convolution model is adjusted with additive spatial random effects and then used to compare the priors for the structured random effect, u_i. An offset variable, *log pop*, of each region *i*, is used as a covariate in the model.

Model: $log\lambda_i = Log\ (pop) + \beta_0 + v_i + u_i;\ v_i \sim iid;\ u_i \sim$ "Besag" ICAR.

Three (3) multilevel models with only one covariate as an offset variable are hereby developed for comparison between the spatial models BYM, the intrinsic CAR and the new "Besag2" ICAR model with additive spatial random effects and given as:

Model 1: $log\lambda_i = Log\ (pop) + \beta_0 + v_i + u_i;\ v_i \sim iid, u_i \sim BYM$

Model 2: $log\lambda_i = Log\ (pop) + \beta_0 + v_i + u_i;\ v_i \sim iid, u_i \sim Besag\ ICAR$

Model 3: $log\lambda_i = Log\ (pop) + \beta_0 + v_i + u_i;\ v_i \sim iid, u_i \sim Besag2\ ICAR$

Comparisons were made between the three spatial dependency models: BYM, ICAR and the newly proposed Besag2 ICAR, to see which of them gave a better fit and more interpretable estimates. As shown in **Table 4.6** below, the major objective is not only to optimize model choice criteria such as DIC values, but to offer a sensible model design where all parameters have a clear significance and interpretation. On the basis of DIC, the performances of the models are almost identical, but the newly proposed model 3 gave the best fit.

Table 4.6: Comparison of Priors for the Structured Random Effects Model without Covariates.

Spatial models	$logpop\ i$	β_0	u_i	v_i	pD	Parameter, a	DIC
BYM	1.24 (0.26)	0.64 (1.37)	2725.85	1.45	23.90	-	**263.12**
ICAR	1.24 (0.26)	0.64 (1.35)	18433.50	1.44	23.90	-	**263.12**
Besag2 ICAR	1.29 (0.27)	0.46 (30.23)	18.83	1.44	23.89	0.05 (0.002)	**263.08**

However, from the reparameterised "Besag2" ICAR model, which has the advantage of possessing two hyperparameters over the Besag ICAR prior with only one hyperparameter, the posterior estimates of the new prior gave significantly reduced variances for the two spatial components, and especially for the structured spatial component ($\sigma = 18.83$), thereby making it to be considered as the best fitted model. In terms of model choice criteria by DIC values, the three models perform at least equally well with existing parameterisations, but only the new model offers parameters that are clearly interpretable.

The comparative spatial maps of the BYM and the usual ICAR models are very identical as shown in **Figure 4.5** below. Spatial maps of ICAR and the new Besag2 ICAR models showed varying disease patterns when the two prior models were compared. The third model which utilized the new prior “Besag2” ICAR showed a better smooth and a more defined disease cluster and distribution.

Areas like Inxuba Yethemba and Camdebbo in Sarah Baatman and Chris Hani municipalities respectively showed varying disease patterns from the two prior models as the new prior gave a smoother and a well-defined spatial dependency.

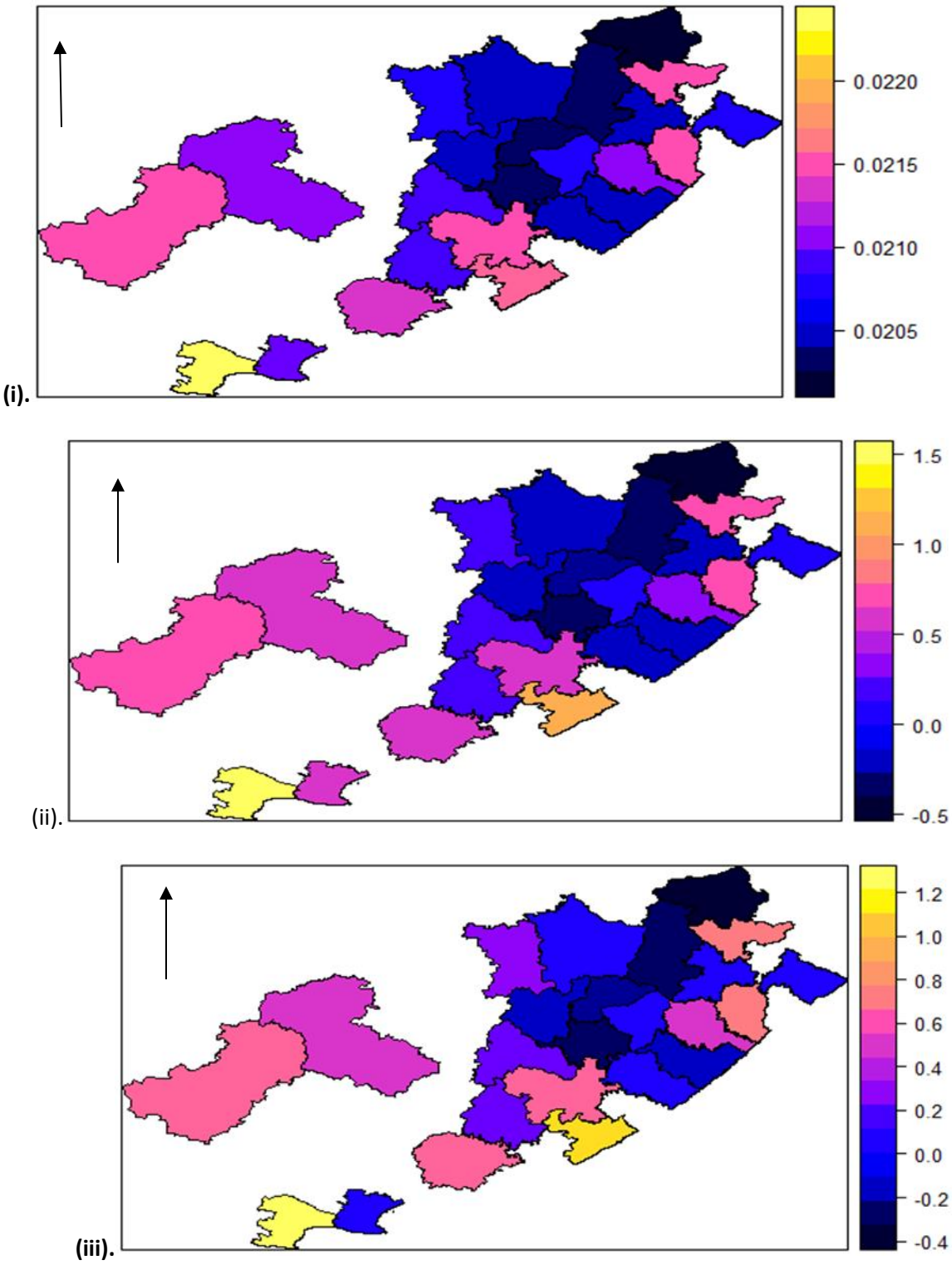


Figure 4.5: Posterior estimated Spatial maps of the structured random prior comparisons of BYM, Besag ICAR and the new "Besag2" ICAR Spatial models respectively.

4.3 Geospatial Assessment

4.3.1. Curve Fittings With Linear And Biharmonic (Spline) Methods For Spatial Patterns Of TB In Eastern Cape, South Africa.

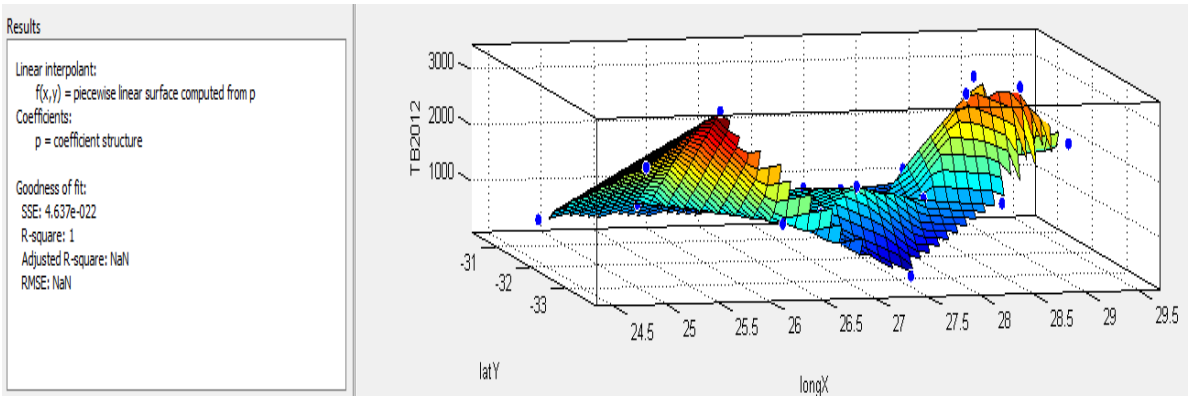
The plottings of all the **Figures of 4.3.1** shown below, displayed the spatial patterns of TB cases by the surface and contour plots using the linear and the biharmonic spline methods for the year (2012-2015). The datasets are represented generally as a 3D or XYZ triplets, where X and Y are the spatial coordinates and Z is the variable of interest and in this case, TB counts in the province.

In curve fitting applications, the interpolant fit category fits an interpolating curve or surface that passes through every data points. Generally, the outlook of all the fittings showed a systematic pattern in the distribution of TB cases in the Eastern Cape province, and this is consistent with some of the initial statistical analyses carried out in this thesis, where the spread of the disease has been found to be non-random. It is also observed that the distribution of TB in the province tends to cluster more at the southern part and also shows a higher disease incidence as it approaches the coastal areas towards the Indian ocean. Only year 2012 showed a two high incidence hotspots of the disease in Kouga (Sarah Baartman Municipality), Nelson Mandela, King Dalingyebo and Nyandeni (OR Tambo) and Umzimvubu in Alfred Nzo municipality. The years 2013, 2014 and 2015 showed similar disease incidence clusters and hotspots.

There is a very high cluster of TB at Amathole, OR Tambo and on the coastal stretch of Sarah Baartman Districts municipalities. Again, they all displayed a worrying pattern of incidence with closer proximity to the Indian ocean.

Neighbourhood effects of the disease spread can also be observed as TB incidence either eases away from the hotspot areas or increase as it approaches the hotspots. The central and northern part of Eastern Cape province showed a very low TB incidence, although not outrightly free from the disease.

(Ai)



A(ii)

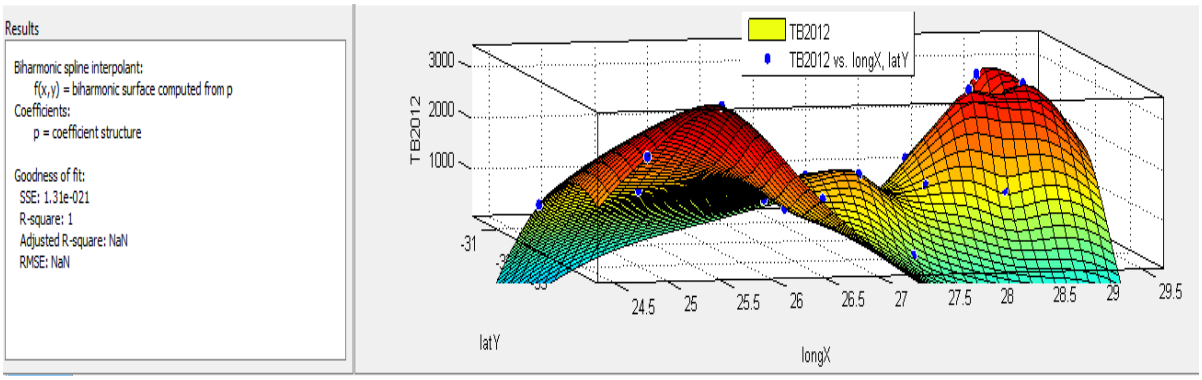
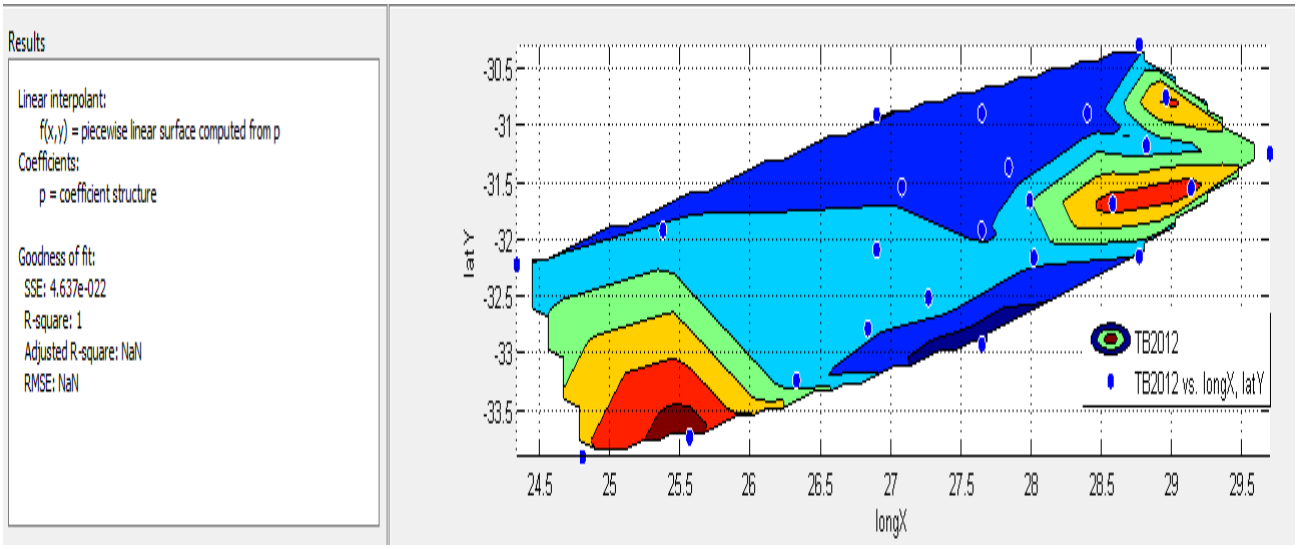


Figure 4.6: Surface plots for A(i) Linear interpolant and (ii) Biharmonic interpolant for TB 2012 data.

B(i)



B(ii)

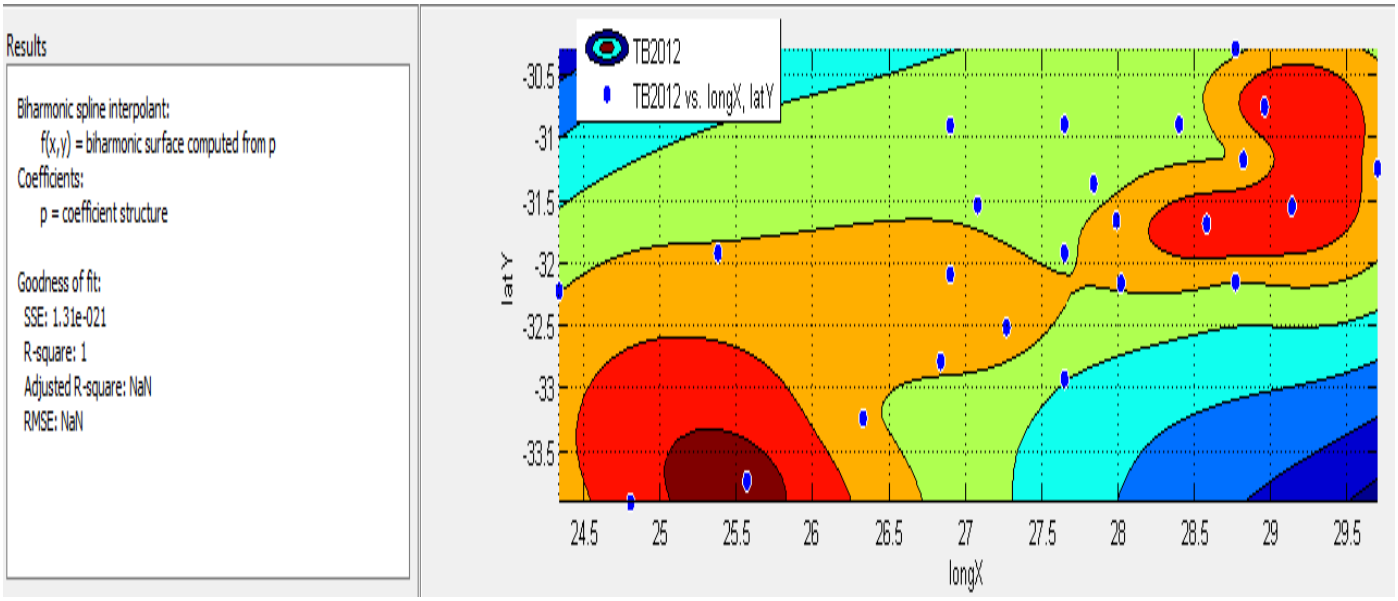
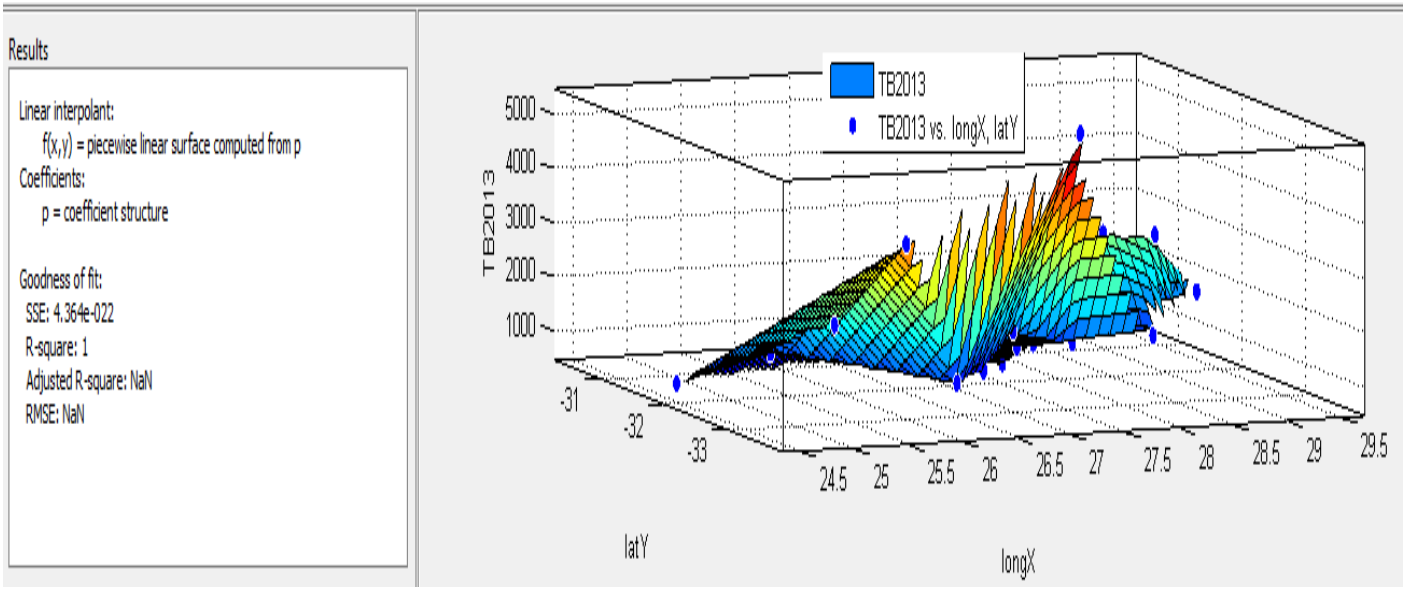


Figure 4.7: Contour plots for B(i). Linear interpolant and B(ii). Biharmonic interpolant for TB 2012 data.

A(i)



A(ii)

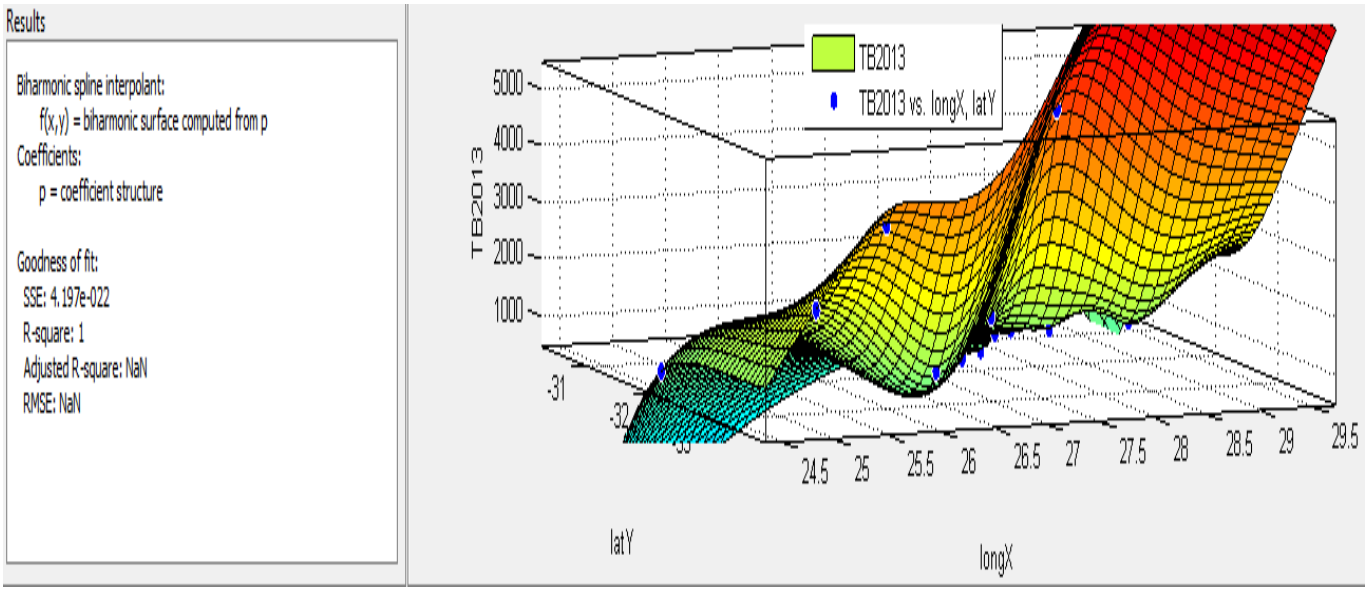
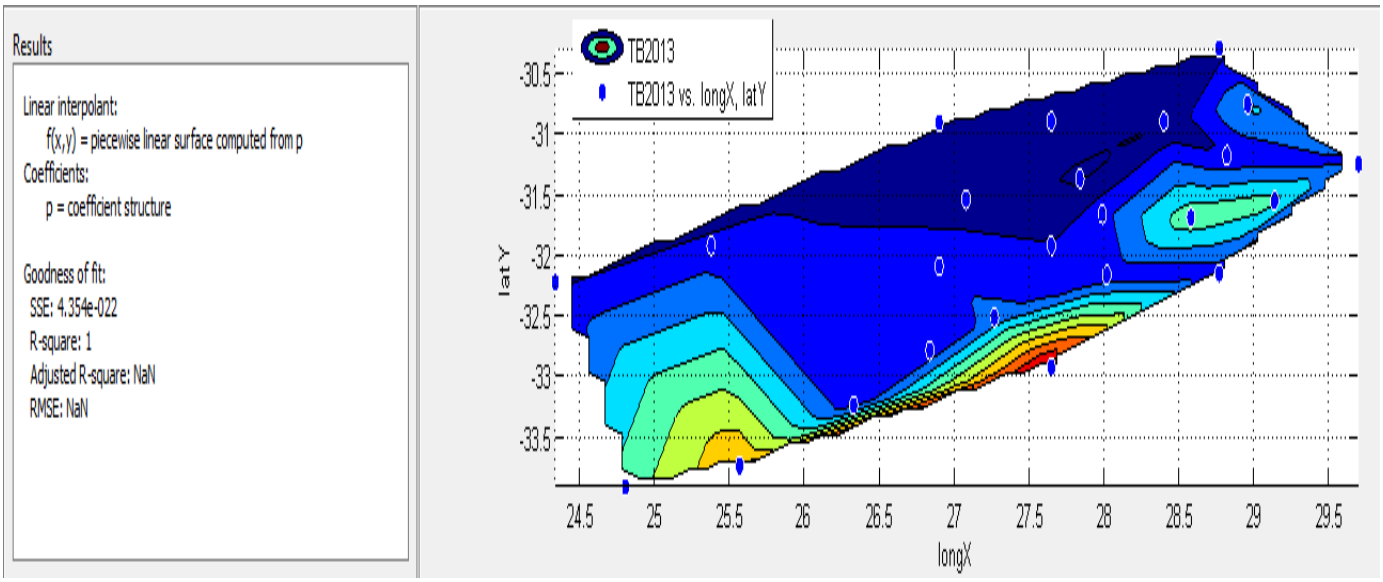


Figure 4.8: Surface plots for A(i) Linear interpolant and (ii) Biharmonic interpolant for TB 2013 data.

B(i).



B(ii)

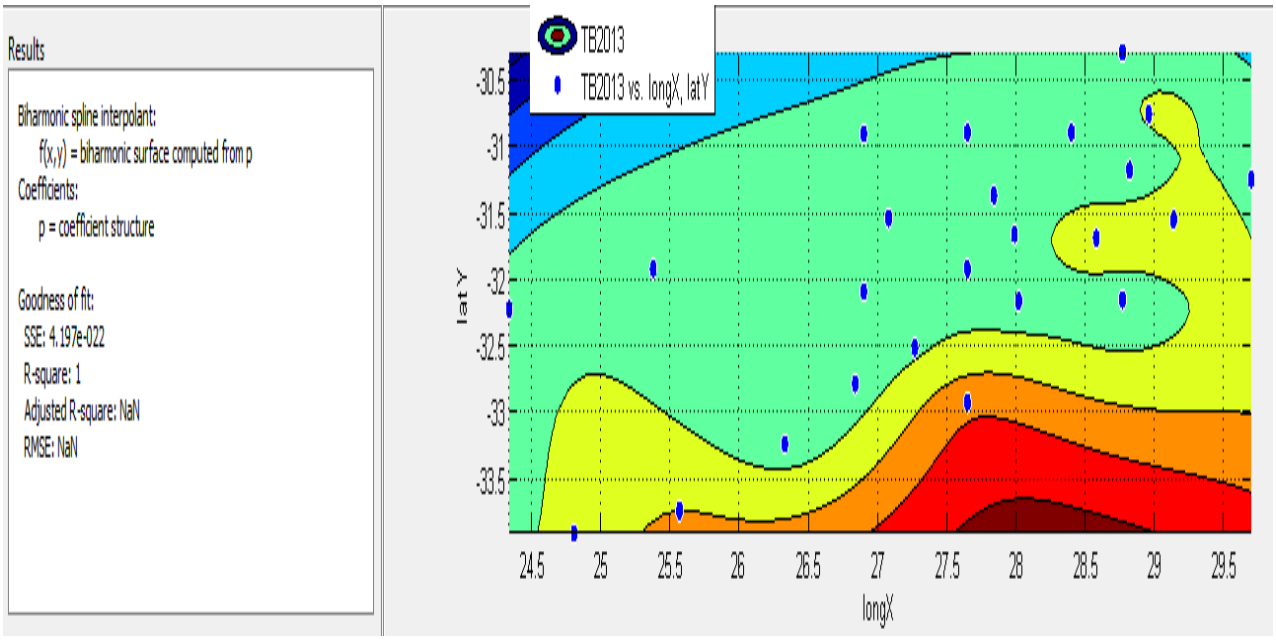
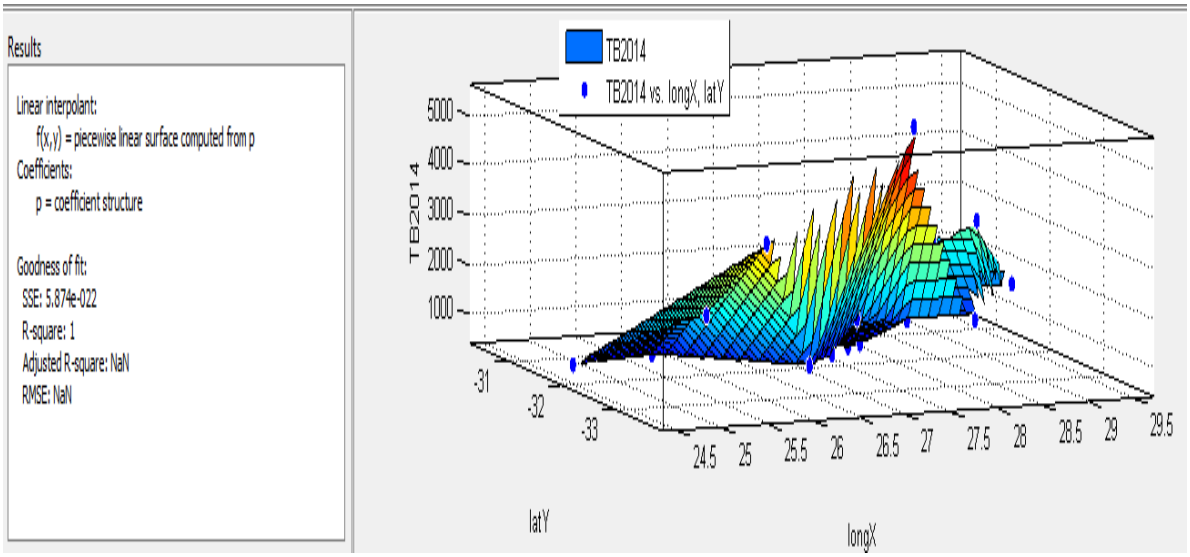


Figure 4.9: Contour plots for B(i). Linear interpolant and B(ii). Biharmonic interpolant for TB 2013 data.

A(i).



A(ii)

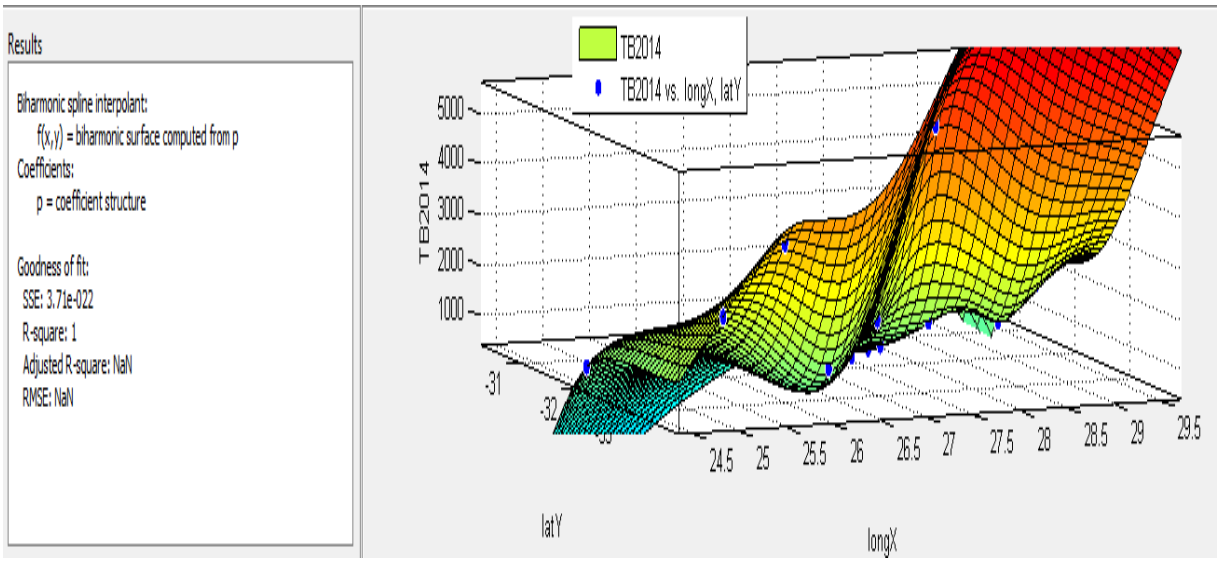
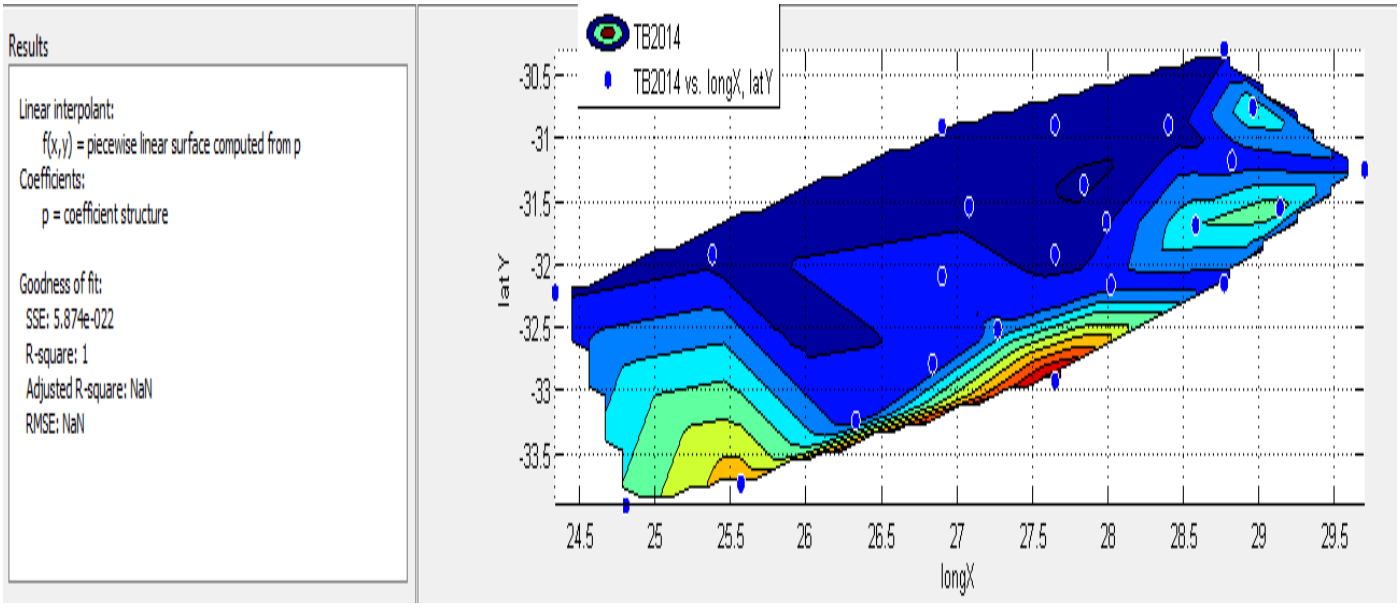


Figure 4.10: Surface plots for A(i) Linear interpolant and A(ii) Biharmonic interpolant for TB 2014 data.

B(i).



B(ii)

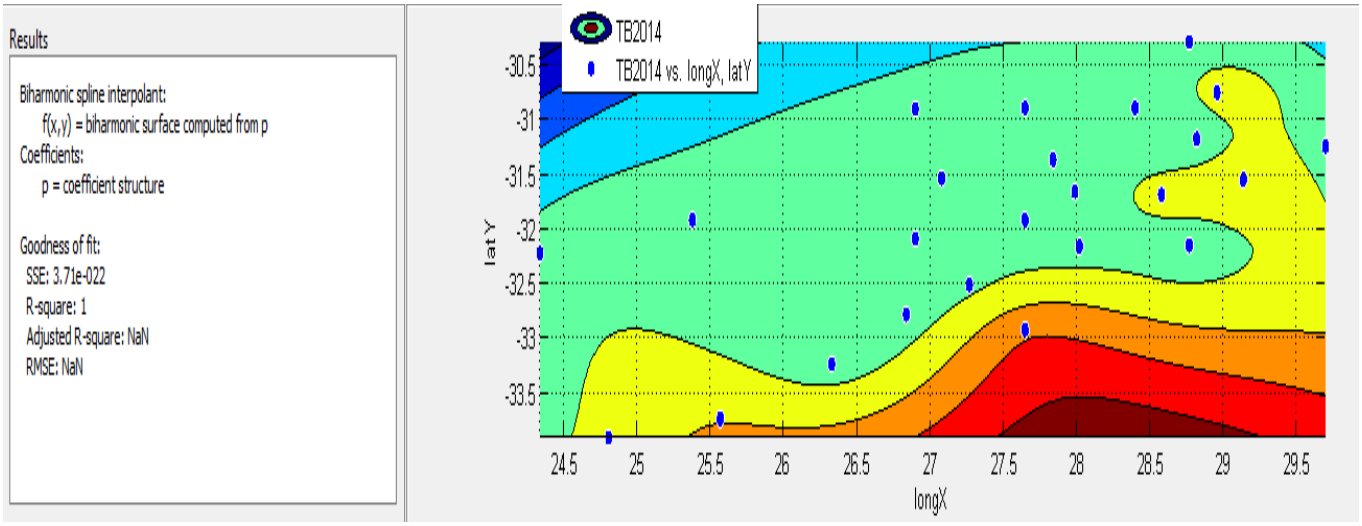
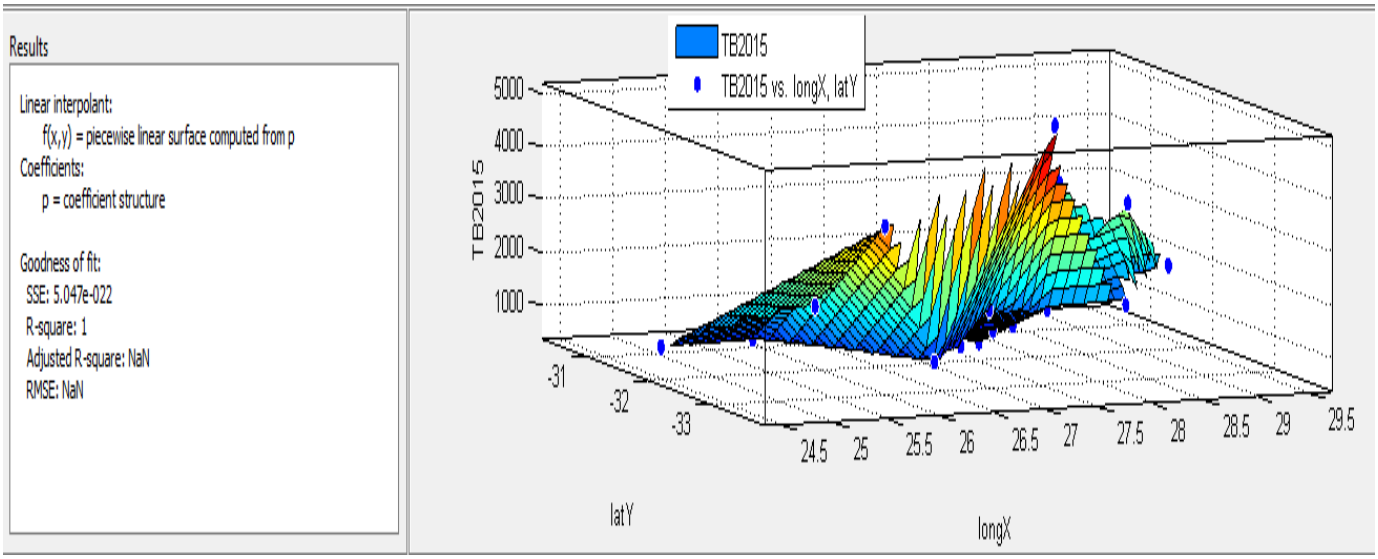


Figure 4.11: Contour plots for B(i). Linear interpolant and B(ii) Biharmonic interpolant for TB 2014 data.

A(i).



A(ii).

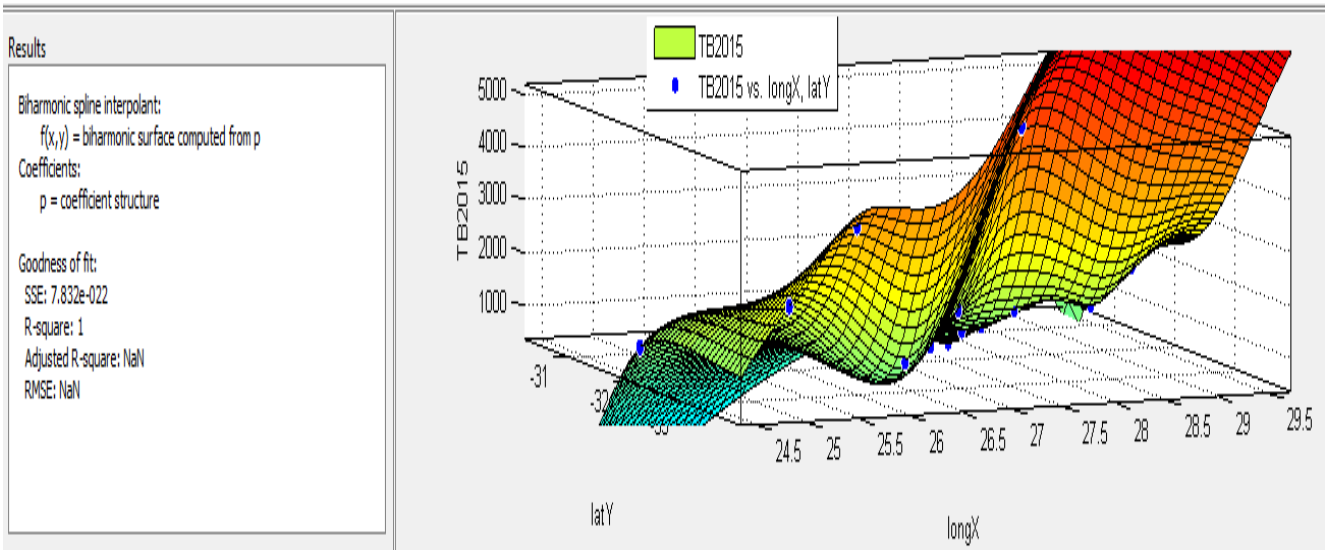
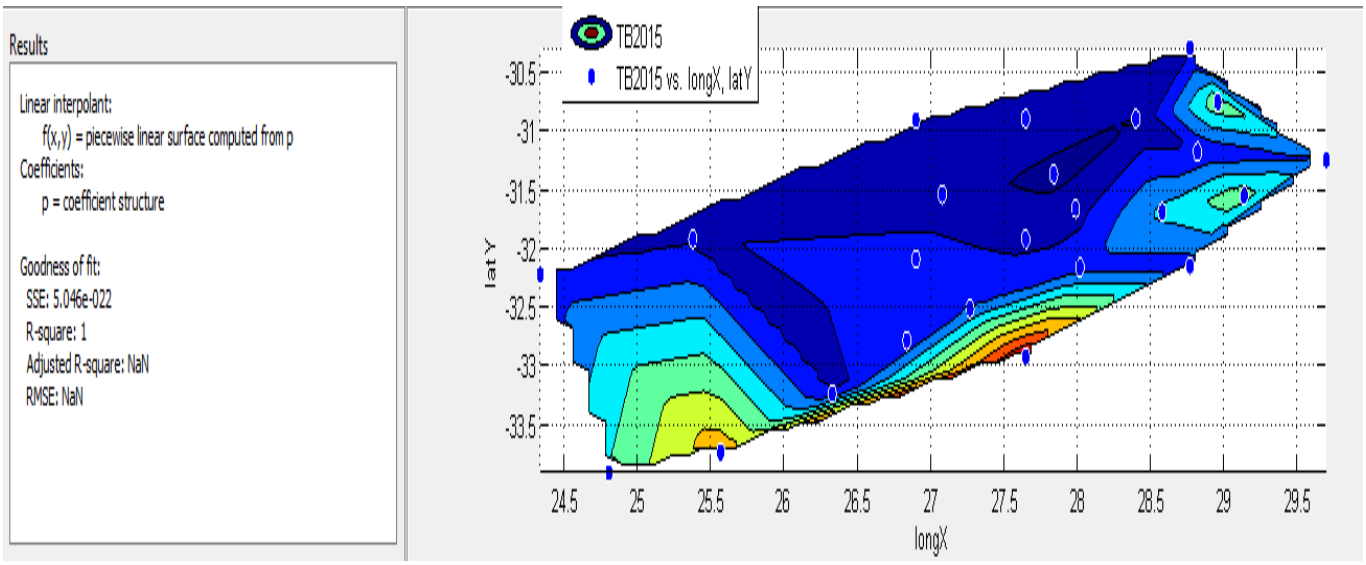


Figure 4.12: Surface plots for A(i). Linear interpolant and (ii) Biharmonic interpolant for TB 2015 data

B(i).



B(ii)

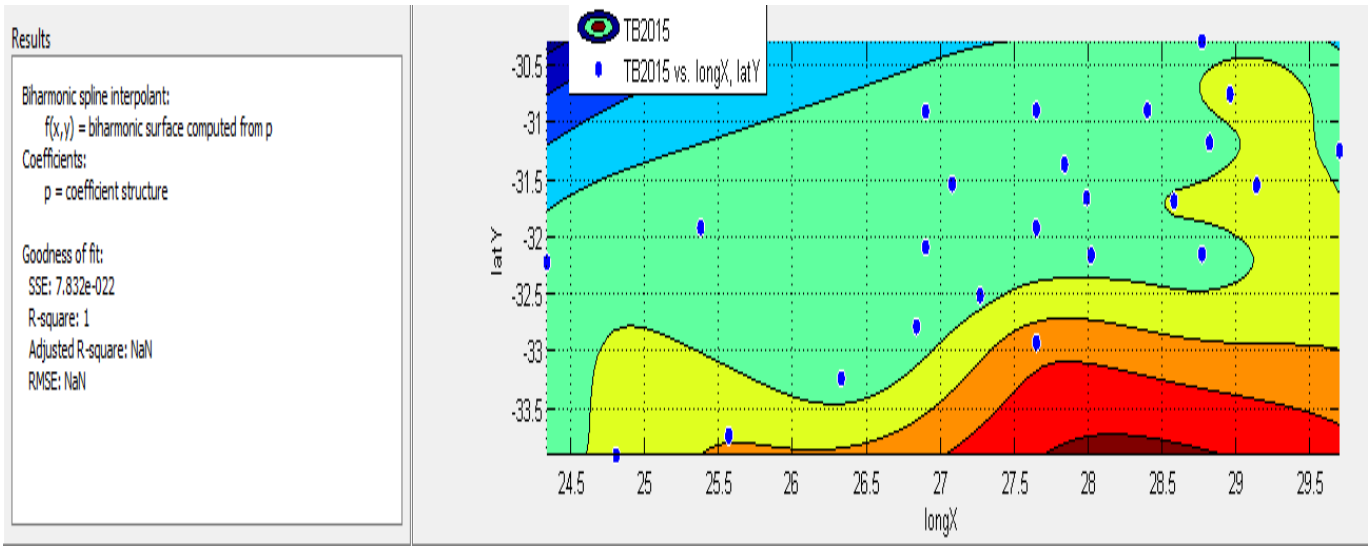


Figure 4.13: Contour plots for B(i). Linear interpolant and B(ii). Biharmonic interpolant for TB 2015 data.

Chapter Five

Findings, Summary and Conclusion

This final chapter summarizes and discuss all the results obtained from the various methods that were employed in the analyses. It also highlights some of the limitations that may be conversant or observed in the whole procedures of this thesis. Reasons and areas of future research are also discussed in this chapter.

5.1 Findings

The results of **Table 4.3** showed the multilevel effects of socio-economic and demographic factors on TB and it was generally observed that the relative risk imposed by poverty, population density and average household size were strong and positively related to the disease outcome. Also, Gini coefficient, unemployment and no schooling had no positive association with TB prevalence in any of the models. In models 3 and 6 where both the proximal and distal factors were combined, showed that some socio-economic factors like Gini coefficient and poverty rates gave both negative and positive relationship in the association respectively. The average number of people living in a household showed a positive association with TB incidence in all the models.

Strong evidences for a relationship between TB and poverty are already available, expressed by higher TB incidence rates in crowed urban areas and amongst low income and illiterate populations (Waaler 2002). It is known that TB spread is cluster dependent, an infected individual in a crowded household would aid the easy transmission of the bacteria as also observed in Souza *et al.*, (2007).

The combination of some socio-economic and demographic factors as distal and proximal factors respectively, considered in this research showed that both risk factors have interplaying effects on TB incidence, as individuals tend to converge with relation to their similar economic and demographic resemblance. The Eastern Cape Province is predominantly Black populated, hence there is a stronger tendency to converge as a community and which makes healthy individuals susceptible to the disease.

The well-known phenomenon of over-dispersion which is inherent with Poisson density in counts data was basically studied in this thesis. This work has presented several regression models for count data allowing for over-dispersion. Over-dispersion is either modelled by the introduction of an additional parameter as in the NB and GP models, or by allowing for an extra proportion of zero observations using zero inflated models or by combining zero inflated models with over-dispersed distributions.

For this thesis, the negative binomial (NB) and the generalized Poisson (GP) distributions were used to capture the extra variation in the TB count data, by adjusting with and without the spatial random components. Usually, additional spatial random effects are included in the models in order to account for unobserved spatial heterogeneity in the data. This method permits for spatial correlations between observations. The result of **Table 4.4** was used for comparison between modelling with Poisson, negative binomial and generalized Poisson for the defined multilevel models. These models were applied to analyse cases of tuberculosis prevalence in the Eastern Cape province of South Africa. The DIC was used for model comparison. The Poisson model gave a significantly best fit than the spatial NB and GP regression models with DIC values of 262.99, 382.87 and 396.14 respectively.

However, the models allowing for over-dispersion (NB and GP) gave better parameter estimates and lower variances for the two spatial random effects when compared with that of Poisson. Among these non-spatial models, the NB model fitted the data best, but the GP model gave more precise estimates from lower variance values.

This thesis showed that even though the Poisson, the Negative Binomial and the Generalized Poisson gave similar estimates for the regression parameters, the standard errors for the Negative Binomial and the Generalized Poisson are higher than the Poisson. Therefore, the Poisson inflated the impact of the regression parameters in the presence of over-dispersion.

The results of **Tables 4.4** and **Table 4.5** have also shown that the choices left for a researcher in choosing a best fit model would be based on either values of DIC or the model with the smaller variance values. Either of these choices have implications but I would suggest for the latter, as it is generally accepted that the lower the variance, the better the precisions.

In concluding, the negative Binomial (NB) and the Generalized Poisson (GP) models are not that difficult to understand. Although the probability density functions for both Negative Binomial and Generalized Poisson involve some mathematically complex formulas, the mean and variance for both models are abstractly easier to be understood. The mean for both Negative Binomial and Generalized Poisson models are equal to the Poisson. The variance of the Negative Binomial is equal or larger than the Poisson, and this allows the Negative Binomial model to handle over-dispersion. The variance of the Generalized Poisson is equal, larger or smaller than the Poisson, and this makes the Generalized Poisson able to handle both over-dispersion and under-dispersion.

A disease mapping technique, which is usually used for smoothening of relative risk is the Bayesian Conditional Autoregressive (CAR) model (Clayton, 1987), which has only one precision hyperparameter, τ. This CAR model offers some shrinkage and spatial smoothing of the raw relative risk estimate, which gives a more stable estimate of the shape of the underlying risk of disease than that given by the raw estimates (Venkatesen *et al.,* 2012).

Without suitable weighting that is characterized with the usual ICAR prior, the hyperparameter usually have no clear meaning and may be incorrectly interpreted. This sole parameter of this ICAR rest on the basic graph structure and is confused with the mixing parameter if the structured spatial effect is not correctly scaled. Also, it is not clear on how to select a prior distribution for this precision parameter. For lack of weighting, a fixed hyperprior for the precision parameter usually give diverse amount of smoothing if the graph on which given disease counts are observed is altered (Riebler *et al.,* 2016). Also, the commonly used hyperprior distributions in the traditional ICAR prior models usually induce overfitting, and will not permit to reduce to simpler models such as a constant risk surface or uncorrelated noise over space (Simpson *et al.,* 2015). Therefore, the need for a weighted and a better latent prior is essential in Bayesian disease mapping.

The spatially structured random effect cannot be treated individually from the unstructured spatial random effect in the classical or frequentist BYM (Besag, York and Mollie) model. This makes the prior explanations for the hyperparameters of the two spatial random effects problematic and challenging (Andrea *et al*., 2016). There are other model designs that address this puzzling issue, nevertheless, the issue on how to select interpretable hyperpriors is still unresolved.

In this thesis, another type of an intrinsic CAR prior, a scaled ICAR prior, which is a member of the latent Gaussian markov random field models with two hyperparameters; a precision parameter,τ , and a weight-parameter, *a*, was developed and assumed for the structured random effect, and compared with the generally used intrinsic conditional autoregressive model ICAR. In this new model, the precision parameter, τ, signifies the marginal precision and it controls the unevenness explained by the spatial effect, while the mixing or weighting parameter, *a*, allocates prevailing variability between the unstructured and structured random effects. This new model parameterizes the BYM model and is also an extension of the Besag ICAR model, that leads to better parameter control as the hyperparameters can be seen independently from each other.

From the result of **Table 4.6**, the new "Besag2" weighted ICAR prior was found to fit the data well with the two spatial components added in the model. The weight-parameter, *a*, of the new prior shows that the new model performed well both showing good learning abilities and good shrinkage behavior. In terms of model choice criteria, the proposed model performed at least equally well as existing parameterisations, but the new "Besag2" prior gave parameters that are interpretable and hyperpriors that have a clear meaning. This parameter meaningfully reduced the variances of the two spatial components, thereby making the new model to be considered as a better fit with a better and more acceptable precisions. Also in terms of model comparison, it was found that the Besag2 model is slightly favored compared to the BYM and the existing ICAR model in terms of DIC in the case of a constant risk.

Performance in terms of model choice criteria by DIC is regarded lesser in this case. The main advantage of this novel "Besag2" ICAR model is that it permits for an intuitive parameter explanation and enables prior assignment. Also, the model is able to shrink towards a spatially unstructured risk for different disease prevalence.

This shows that the Besag2 ICAR model does not overfit the parameter estimates. It is therefore acceptable, that the practical advantages in terms of interpretability and prior assignment makes the newly proposed Besag2 ICAR model in this thesis beneficial compared to existing models, and its usage is also recommended since its model criteria performance is at least as good as for existing approaches.

Surface modeling is the process of a natural or artificial surface determination by using one or more mathematical equations. Modeling the 3-dimensional surface in space involves finding a function z = f (x, y) that represents the entire surface of the values z = f (x, y) associated with the point P (x, y) arranged irregularly. In addition, this function can predict the values z = f (x, y) and for other positions regularly arranged. Such a feature is known as interpolation function (Dimitru *et al.,* 2013). Interpolation and gridding of data are processes in the physical sciences and are accomplished naturally using an averaging or finite difference scheme on an equidistant grid. Cubic splines are common because of their smooth appearances however, these functions can have undesirable oscillations between data points (Wessel *et al.*, 1998).

In linear methods, the new surface is restricted to the area that contains the disease points and by default, it does not extrapolate beyond its regions. Also, the contours in linear method are rather angular than when compared with the biharmonic methods, which smooth the shapes of the contours. The method also presents a fast interpolation algorithm but has the disadvantage that the interpolation function is limited to the area bounded by convex random set of data points, thereby resulting in a surface that is not smoothed. MATLAB offers a biharmonic spline interpolation since the early stages. This interpolation method was developed by Sandwell (1987). This specific gridding method produces smooth surfaces that are mainly appropriate for noisy data sets with irregular distribution of control points.

On the observed TB counts from the twenty four (24) health subdistricts of the Eastern Cape province, we used the linear and the biharmonic spline interpolation methods for the disease spatial distribution. These two methods were used for comparison both for the surface and contour plottings, and the latter (biharmonic) was observed to give a well defined disease map and fit from the values of their respective sum of squares errors (SSE), and it also displayed a better spatial pattern of TB prevalence for the regions. On the biharmonic smoothing method, the spline interpolation is used to interpolate the irregular-spaced dataset at grid points.

The biharmonic spline interpolation method used in this thesis provides a solution to most gridding problems. The biharmonic spline interpolations are in many ways, said to be the other extreme, because they can be used for very irregular-spaced and noisy data, hence the contours suggest an extremely smooth surface. For reasonable amounts of data, the Green's function or the biharmonic technique is superior or higher to the orthodox finite-difference methods because (1) both data values and directional gradients can be used to constrain the model surface, (2) noise and disturbances can be curbed easily by seeking a least-squares fit rather than exact interpolation, and (3) the model can be evaluated at random locations rather than only on a rectangular grid. In conclusion, one of the most worthwhile features of the biharmonic function approach for splines lies in the great simplification of the computer implementation and interpretation of the method.

5.2 Conclusion

This thesis has been concerned with developing and extending statistical models in the area of spatial modelling with application to Tuberculosis data in the Eastern Cape Province of South Africa. The model (BYM) under consideration provides for areal (lattice) data only. Point pattern data were not examined in this work.

Earlier disease mapping models are based on collating, mapping and analysing prevalence or incidence data with conventional (frequentist) statistical methods, which are often affected by random variation due to population variability and a loss of statistical power when cases are given to subgroups in geographical subregions.

The true spatial distribution of a disease may not be reflected by the observed extreme values; instead it reflects those of the population areas. The Bayesian approach can overcome these problems as it can model the random and true variation separately (Bergamaschi *et al.*, 2006). This is the basis upon which this thesis is centered on and it is an attractive substitute to the frequentist approach.

This research by extension, considered an additional factor of demographic variability to socio-economic factors as a possible risk factor for the prevalence of TB in the Eastern Cape province of South Africa. A combination of a Bayesian and generalized linear mixed model (GLMM) parameter estimation was carried out using the integrated nested Laplace approximation (INLA) and implemented in R, as an alternative to the commonly used Monte Carlo Markov Chain (MCMC) simulation technique. The former was adopted because of its computational advantage of speed over MCMC.

The practical implications of the multilevel models developed and utilised in this thesis to public health, are in its ability to direct for areas of surveillance and interventions to these disease hotspots. The spatial patterns of the disease distribution, as seen from the disease maps, would serve as a pointer to epidemiologists and government health agencies, to be able to make policies that would curb the spread of the disease and also promote health education to the concerned population of those hotspots. Areas around the coastal lines, which showed a higher disease prevalence from the spatial maps, should be given higher priority, and multifaceted factors that could be responsible for such a trend should be investigated, and those concerned should be properly educated.

The struggle against infectious diseases like tuberculosis, will not decrease in the predictable future, so biostatisticians and epidemic modelers must continue to advance our science in order to contribute meaningfully to the fight. Also, government and public health authorities should intensify efforts in and for the Eastern Cape Province, to ensure that regions and locations with a consistent rise in TB prevalence are specially targeted with appropriate interventions. When carrying out the various preventive interventions, emphasis should be focused also on preventing the incidence of HIV as it has a positive alliance with tuberculosis.

5.3 Limitations of the Study and Future Research

As we know that no study goes without limitations, a major limitation of our study is that data on sex and age of the cases were not provided by the electronic tuberculosis register (ETR), hence we could not give any case description for these groups.

For this thesis, future works would consider extensions of the models presented to model for the negative binomial distribution, by allowing the dispersion parameter, d, to be spatially correlated and not fixed as assumed.

Further extensions of mixed negative binomial distribution that can cater for both under-dispersion and over-dispersion have been mentioned for count data analyses (Tomoaki Imoto, 2014).

Further studies can also be directed into using alternative distributions for the structured random effects in lieu of the intrinsic conditional autoregressive (ICAR) model. The Besag2 ICAR model can also be logically joined and applied with covariate information, or integrated into a spatio-temporal situation. Further work would also be required to distribute the variance not only within the spatial components but over all model parameters in the linear predictor.

A major extension of all the models discussed in this thesis is the incorporation of the temporal component as most spatial data are also time dependent. Since ETR survey will be carried out in the future, these models can be extended to accommodate for this time portion. Some authors have considered temporal extensions to hierarchical spatial models based on a parametric description of the time trends, on independent risk estimates for each time period, or on the definition of the joint covariance matrix for all the periods as a Kronecker product of matrices (Cressie and Wikle, 2011).

Nonparametric methods can also be employed in the overall analyses of Bayesian spatial modelling. A full semi-parametric or penalized regression based on splines is an area to look into, as it would relax the restrictions involved and embedded in parametric procedures.

Above all, further researches can be focused on extending these models for TB/HIV joint modelling, and this should be given a high priority research concern as it is a major problem in South Africa. Research on the distribution of tuberculosis can also be extended into the racial composition of the province and the country in general, as South Africa still has evidence of racial divide by residential differentiation in some locations.

Bibliography

ABS (2006). Australian Standard Geographical Classification. *Technical Report*, ABS Cat. No. 1216.0. Australian Bureau of Statistics, Canberra.

Agresti, A. (2002). Categorical Data Analysis. *Wiley*, Second Edition.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. 2nd Internat. Sympos.Inform. *Theory, Tsahkadsor* 1971, 267-281.

Allison, P. D. (1996). Fixed Effects Partial Likelihood For Repeated Events. *Sociological Methods & Research* 25: 207-222.

Allison, P. D. (1999). Logistic Regression Using the SAS System: Theory and Applications. Cary, NC: *SAS Institute*.

American Thoracic Society/CDC. (1994). Treatment of Tuberculosis and Tuberculosis Infection in Adults and Children. *Am J Respir Crit Care Med* 1994; 149:1359-1374.

Anderson, C., Lee, D. And Dean, N. (2014). Identifying Clusters in Bayesian Disease Mapping. *Biostatistics* 15, 457–469.

Anderson, R. M. and May, R. M. (1991). Infectious Diseases of Humans: Dynamics and Control, *Oxford University Press*.

Ando, A. And Kaufman, G. (1965). Bayesian Analysis of the Independent Multinomial Process-Neither Mean nor Precision Known. *Journal of the American Statistical Association* (In Press) Vol. 60, No. 309, Pp. 347-58.

Andrew, B. L. And David, G. T. (2002). Spatial Cluster Modelling. *Chapman and Hall/CRC*. ISBN 9781584882664 - CAT# C2662.

Anscombe, F.J. (1950). Sampling Theory of Negative Binomial and Logarithmic Series Distributions. *Biometrika*, 37, 358-382.

Anselin L. Local Indicators Of Spatial Association-LISA *Georg Anal*. (1995). 27:91-115.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. *Kluwer Academic Publishers*, Dordrecht.

Anselin, L. (2001). Spatial Econometrics. In Baltagi, B. (Ed.), A Companion To Theoretical Econometrics. *Oxford: Basil Blackwell*, Pp. 310-330.

ATS/CDC Statement Committee on Latent Tuberculosis Infection (June 2000). Targeted Tuberculin Testing and Treatment of Latent Tuberculosis Infection. *American Thoracic Society*. MMWR. Recommendations and Reports **49** (RR–6): 1–51. PMID 10881762.

Audit Scotland (2012). Health Inequalities in Scotland. Technical Report, the Scottish Government.

Bailey, T C, Carvalho, M.S, Lapa, T.M, Souza, W.V. and Brewer, M.J. (2005). Modelling Of Under-Detection in Disease Surveillance. *Annals of Epidemiology* 15, 335-343.

Bailey, T. C. (2001). Spatial Statistiscal Methods in Health. *Cadernos De Saude Publica* 17, 1083-1098.

Ball, F. D. Mollison and G. Scalia-Tomba (1997). Epidemics with Two Levels of Mixing. *The Annals of Applied Probability* 7: 46-89.

Banerjee, S., Carlin, B.P. And A.E. And Gelfand. (2004). Hierarchical Modelling and Analysis for Spatial Data. *Chapman and Hall/CRC*, Florida, U.S.A.

Barr, R.G., Diez-Roux, A.V., Knirsch, C.A. and Pablos-Méndez, A. (2001). Neighborhood Poverty and the Resurgence of Tuberculosis in New York City, 1984–1992. *Am J Public Health*. 91:1487–1493.

Bell, B. And Broemeling, L. (2000). A Bayesian Analysis of Spatial Processes with Application to Disease Mapping. *Stat. Med*. 19, 957-974.

Bergamaschi, R., Montomoli, C., Candeloro, E., Monti, M.C., Cioccale, R., Bernadinelli, L., Fratino, P. and Cosi, V. (2006). Bayesian Mapping of Multiple Sclerosis Prevalence in the Province Of Pavia, Northern Italy. *J Neurol Sci.*, 244(1-2): 127-131.

Bernardinelli, L., D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini. (1995). Bayesian Analysis of Space-Time Variation in Disease Risk. *Statistics in Medicine* 14, 2433-2443.

Besag J, York J, Mollie´A. (1991). Bayesian Image Restoration, With Two Applications in Spatial Statistics. *Annals of the Institute of Statistical Mathematics*; 43: 1–59 (With Discussion).

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*; 36: 192–236 (With Discussion).

Besag, J. (1981). On A System of Two-Dimensional Recurrence Equations. *Journal of the Royal Statistical Society Series B* (Statistical Methodology); 43: 302–9.

Best, N., Arnold, R., Thomas, A., Waller, L. And Conlon, E. (1999). Bayesian Models for Spatially Correlated Disease and Exposure Data. In: Bernaddo, J. *Et Al.* (Eds.). Disease Mapping and Risk Assessment for Public Health, *Chichester: Wiley*.

Best, N., Richardson, S. And Thompson, A. (2005). A Comparison of Bayesian Methods for Disease Mapping. *Statistical Methods in Medical Research* 14, 2433-2443.

Best, N.G, Ickstadt, K. and Wolpert, R.J. (2000). Spatial Poisson Regression For Health And Exposure Data Measured At Disparate Resolutions. *Journal of the American Statistical Association*; 95: 1076–1088.

Best, N.G, Ickstadt, K., Wolpert, R.L., Cockings, S., Elliott, P., Bennett, J., Bottle, A. and Reed, S. (2002). Modelling the Impact of Traffic-Related Air Pollution on Childhood Respiratory Illness. In Gatsonis C, Kass RE, Carlin B, Carriquiry A, Gelman A, Verninelli I, West M Eds. Case Studies In Bayesian Statistics Volume V. New York, NY, USA: *Springer-Verlag*: 183–259 (With Discussion).

Best, N.G., Ickstad, K., Wolpert, R.L. and Briggs, D.J. (2000). Combining Models of Health and Exposure Data: The SAVIAH Study. In Elliott P, Wakefield JC, Best NG, Briggs DJ Eds. Spatial Epidemiology: Methods and Applications. Oxford: *Oxford University Press*: 393– 414.

Bishai, W.R., Graham, N.M. and Harrington, S. (1998). Molecular and Geographical Patterns of Tuberculosis Transmission after 15 Years of Directly Observed Therapy. *JAMA*. 280:1679-1684.

Bithell, J.F. (2000). A Classification of Disease Mapping Methods. *Statistics in Medicine*. 19(17-18): 2203-2215.

Bjornstad, O.N., B. F. Finkenstadt and B. T. Grenfell (2002). Dynamics of Measles Epidemics: Estimating Scaling Of Transmission Rates Using a Time Series SIR Model. *Ecological Monographs* 72: 169-184.

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Lutchter, S. and Spicer, R. (2000). The Economic Impact of Motor Vehicle Crashes. Report No. DOT HS 809 446, *National Highway Traffic safety Administration*, Washington, D.C.

Bliss, C.I. and Fisher, R.A. (1953). Fitting the Negative Binomial Distribution to Biological Data. *Biometrics* 9: 176–200.

Bliss, C.I. and Owen, A.R.G. (1958). Negative Binomial Distributions with a Common K. *Biometrika* 45: 37-58.

Blum, H. F. (1948). Sunlight as a Causal Factor in Cancer of the Skin of Man. *Journal of the National Cancer Institute*, 9:247-58.

Bohning, D. (2003). Empirical Bayes Estimators and Non-Parametric Mixture Models for Space and Time Space Disease Mapping and Surveillance. *Environmetrics* 14, 431-451.

Bor, D. H., and Epstein, P. R. (1991). Pathogenesis of Respiratory Infection in the Disadvantaged. *Semin. Respir. Infect*. 6, 194±203.

Box, G. E. P., and Tiao, G. C. (1973). Bayesian Inference in Statistical Analysis. Reading, Mass.: *Addison-Wesley*.

Bradley Efron and Carl Morris. (1975). Data Analysis Using Stein's Estimator and Its Generalisations. *Journal of the American Statistical Association*. Vol. 70, No. 350.Pp 311-319.

Bradshaw, D., Groenewald, P., Laubscher, R., Nannan, N., Nojilana, B., Norman, R., Pieterse, D. and Schneider, M. (2003). Initial Estimates from the South African National Burden of Disease Study, 2000. *MRC Report*. ISBN: 1-919809-64-3.

Breslow, N, E. and G. Clayton, D. (1993). Approximate Inference In Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88, 9-25.

Brody H, Rip MR, Vinten-Johanson P, Paneth N, and Rachman, S. (2000). Map-Making and Myth-Making in Broad Street: *The London Cholera Epidemic,* 1854. Lancet; 356; 64-69.

Cameron, A. Colin and Pravin K. Trivedi. (1998). Regression Analysis of Count Data. *Cambridge, UK: Cambridge University Press*.

Cameron, A.C., & Trivedi, P.K. (1998). Regression Analysis of Count Data. New York: *Cambridge University Press*.

Carl N. Morris. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*. Vol. 78, No. 381 Pp. 47-55.

Carlin B P. And T A. Louis. (2000). Bayes and Empirical Bayes Methods for Data Analysis. Boca Raton: *Chapman/CRC Press*.

Casetti, E. (1992). Bayesian Regression and the Expansion Method. *Geograp. Anal*. 24; 58-74.

Chib, S. (1993). Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach. *Journal of Econometrics* 58; 275-294.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *J. Am. Stat. Assoc*. 90, 1313-1321.

Clark, S.J. and Perry, J.N. (1989). Estimation of the Negative Binomial Parameter Kappa by Maximum Quasi-Likelihood. *Biometrics* 45: 309–316.

Clayton, D. And Kaldor, J. (1987). Empirical Bayes Estimate of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics*. 43:671-681.

Clayton, D. G. And Bernardinelli, L. (1992). Bayesian Methods for Mapping Disease Risk. In: Elliott, P., Cuzick, J., English, D. and Stern, R. (Editors), Geographical and Environmental Epidemiology: Methods for Small Area Studies. Oxford: *Oxford University Press*. Pp 205–210.

Congdon, P. (2001). Bayesian Statistical Modelling. New York: *Wiley*.

Congdon, P. And Southall, H. (2005). Trends in Inequality in Infant Mortality in the North of England, 1921 To 1973, And Their Association with Urban and Social Structure. *Journal of the Royal Statistical Society: Series A.* 168; 679-700.

Congdon, Peter. (2003). Applied Bayesian Modelling. *Wiley Series in Probability and Statistics*. ISBN 0-471-48695-7.

Cook, D.G. And Pocock, S. J. (1983). Multiple Regression in Geographic Mortality Studies with Allowance for Spatially Correlated Errors. *Biometrics* .39: 361–71.

Cook, J.D. (2009). Notes on the Negative Binomial Distribution. www.Johnd Cook.Com/Negative Binomial.Pdf (Accessed on Aug. 15th, 2010).

Core R Development. R. (2005). A Language and Environment for Statistical Computing.

Cressie, N., Perrin, O. And Thomas-Agnan, C. (2005). Likelihood-Based Estimation for Gaussian Mrfs. *Statistical Methodology*, 2, 1–16.

Cressie, N.A. And Chan, N. (1989). Spatial Modelling Of Regional Variables. *Journal of the American Statistical Association*, 84 (406), 393-401.

Cressie. N.A.C. (1990). The Origins of Kriging. *Mathematical Geology*, Vol. 22, Pg. 239-252.

Daiane Leite Da Roza, Maria Do Carmo Gullaci Guimarães Caccia-Bava and Edson Zangiacomi Martinez. (2012). Spatio-Temporal Patterns Of Tuberculosis Incidence In Ribeirão Preto, State Of São Paulo, Southeast Brazil, And Their Relationship With Social Vulnerability: A Bayesian Analysis. *Revista Da Sociedade Brasileira De Medicina Tropical* 45(5):607-615.

De Graaff, T., Florax, R., Nijkamp, P. And Reggiani, A. (2001). A Generalized Misspecification Test for Spatial Regression Models: Dependence, Heterogeneity and Nonlinearity. *J. Regional Sci*. 41, 255-276.

Denison, D.G.T, Holmes, C.C, Mallick, B.K. And Smith, A.F.M. (2002). Bayesian Methods for Nonlinear Classification and Regression. New York, NY, USA: *John Wiley & Sons*.

Denison, D.G.T,.And Holmes, C.C. (2001). Bayesian Partitioning For Estimating Disease Risk. *Biometrics*. 57:143–9.

Department Of Health, (Doh) South Africa (2012). Annual Performance Plan 2012/13 – 2014/15. (See More At: Http://Www.Tbfacts.Org/Tb-Statistics-South-Africa).

Diekmann, O., Heesterbeek, J. A. P. And Metz, J.A.J. (1990). On The Definition and the Computation of the Basic Reproduction Ratio R_0 in Models for Infectious Diseases in Heterogeneous Populations. *Journal of Mathematical Biology* 28: 365-382.

Dietz, K. And Hadeler, K.P. (1988). Epidemiological Models for Sexually-Transmitted Diseases. *Journal of Mathematical Biology* 26: 1-25.

Diggle, P. J. (1983). Statistical Analysis of Spatial Point Patterns. Mathematics in Biology, Vol. 2. *Academic-Press*, London - New York. 148 S.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1988). Model-Based Geostatistics. *Journal of the Royal Statistical Society Series C* (Applied Statistics) 47: 299–350 (With Discussion).

Dolores, A., Kimberly A.L., Theresa, L.O. And Subramanian,S.V. (2003). Future Directions in Residential Segregation and Health Research: A Multilevel Approach. *Am J Public Health.* 93:215–221.

Dormaan, C.F. (2007). Effects of Incorporating Spatial Autocorrelation into the Analysis of Species Distribution Data. *Glob. Ecol. Biogeogr*., 16, 129-138.

Eastern Cape AIDS Council (2012). Provincial Strategic Plan for HIV and AIDS, STDs and TB 2012-2016.

Eberly, L.E. And Carlin, B.P. (2000). Identifiability and Convergence Issues for Markov Chain Monte Carlo Fitting Of Spatial Models. *Statistics in Medicine.* 19: 2279–2294.

Elliott, P. And Wartenberg, D. (2004). Spatial Epidemiology: Current Approaches and Future Challenges. *Environ Health Perspect* 112:998–1006.

Elliott, P., Wakefield, J.C., Best, N.G. And Briggs, D.J. Eds. (2000). Spatial Epidemiology: Methods and Applications. Oxford: *Oxford Univ. Press*.

Fotherimgham, A., Brunsdon, C. And Charlton, M. (2000). Quantitative Geography. *London. Sage*.

Gaetan, C. And Guyon, X. (2010). Spatial Statistics and Modelling, Springer series In Statistics, XIV, 302 p.

Gandin, L.S. (1963). Objective Analysis of Meteorological Fields. Gidro-Meteorologicheskoe Izdatelstvo (GIMIZ), Leningrad. (Israel Program For Scientific Translations, Jerusalem, 1965, Pg 242).

Gangnon, R.E, Clayton, M.K. (2000). Bayesian Detection and Modelling Of Spatial Disease Clustering. *Biometrics*. 56: 922–35.

Gani, R. And Leach, S. (2004). Epidemiologic Determinants for Modelling Pneumonic Plague Outbreaks. *Emerging Infectious Diseases* 10: 608-614.

Gelfand, A. And Ghosh, S. (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika* 85(1): 1-11.

Gelfand, A. And Ghosh, S. (2000). Generalized Linear Models: A Bayesian View. In: Dey, D., Ghosh, S. and Mallick, B. (Eds.), Generalised Linear Models: A Bayesian Perspective. *New York: Merkel-Dekker*, Pp. 1-22.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85(410): 398-409.

Gelfand, A.E. And Vounatsou, P. (2003). Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis, *Oxford University Press.* Biostatistics 41; Pp. 11–25.

Gelfand, A.E., Hills, S.E., Racine-Poon, A. And Smith, A.F.M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *J. Amer. Statist. Assoc*. 85, 972–985.

Gelman, A., Carlin, J. B., Stern, H. S. And Rubin, D. B. (2003). Bayesian Data Analysis, Second Edition. *London: Chapman and Hall*.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). Bayesian Data Analysis. *Chapman & Hall/CRC*, Second Edition.

Gelman, A., J. Hwang, and Vehtari, A. (2014). Understanding Predictive Information Criteria for Bayesian Models. *Statistics and Computing* 24, 997-1016.

Gelman, Andrew (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis* 1, Number 3, Pp. 515–533.

Gemperli, A. (2003). Development Of Spatial Statistical Methods For Modelling Point Referenced Spatial Data In Malaria Epidemiology. Basel: University Of Basel. (Unpublished, Phd Thesis).

Ghosh, M. And Rao, J. (1994). Small Area Estimation: An Appraisal. *Stat. Sci*. 9, 55-76.

Grassberger, P. (1983). Critical Behavior of the General Epidemic Process and Dynamical Percolation. *Math Biosci*. 63(2):157-172.

Green, P. And Richardson, S. (2002). Hidden Markov Models and Disease Mapping. *Journal of the American Statistical Association*. 97:1055–70.

Grenfell, B. T. Bjornstad, O.N. And Kappey, J. (2001). Travelling Waves and Spatial Hierarchies in Measles Epidemics. *Nature* 414: 716-723.

Haining, R.P. (1990). Spatial Data Analysis in the Social and Environmental Sciences. *Cambridge University Press*, Cambridge.

Haining, R.P. (1994). Designing Spatial Data Analysis Modules for Geographical Information Systems. Pages 45-64 In Fortheringham S. And Rogerso P. (Eds) *Spatial Analysis and GIS, Taylor and Francis*, London.

Haining, R.P. (2003). Spatial Data Analysis: Theory and Practice. *Cambridge University Press*, Cambridge.

Harling, G. And Castro, M.C. (2013). A Spatial Analysis of Social and Economic Determinants of Tuberculosis in Brazil. *Health Place*. Vol. 25:56-67.

Harrison, Hong and Stein, J.C. (1999). A Unified Theory Of Under Reaction, Momentum Trading, and Overreaction in Asset Markets. *The Journal of Finance*, Vol. LIV(6): 2143-2184.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 57, 97-109.

Hetherington, H.W., Landis, M. And Opie, A. (1929). Survey to Determine the Prevalence of Tuberculosis Infection in School Children. *Am Rev Tuberc*. 1929:421.

Higdon, D.M. (2002). Space and Space-Time Modelling Using Process Convolutions. In Anderson CW, Barnett V, Chatwin PC, El-Shaarawi AH Eds. *Quantitative Methods for Current Environmental Issues*. *London, UK: Springer-Verlag*: 37–54.

Hoge, C. W., Fisher, L., Donnell Jr, H. D., Dodson, D. R.,Tomlinson Jr, G. V., Breiman, R. F., Bloch, A. B. And Good, R. C. (1994a). Response To `Invited Commentary: Relative Susceptibility of Black Americans to Tuberculosis'. *Am. J. Epidemiol.*, 139; 533-534.

House, J. S., Landis. K.R., And Umberson, D. (1988). Social Relationships and Health. Science 241: 540-545.

Ita Kreft And Lan De Leeuw. (1998). Introduction to Multilevel Modelling. *Sage Publications Ltd*. London RC2A 4PU. ISBN 0 7619 5140 7.

Jackman, S. (2000). Estimation and Inferences "Missing Data" Problems: Unifying Social Science Statistics via Bayesian Simulation. *Political Analysis*, 8(4), 307-322.

James, S.H. And Brian, J.R. (2010). Adding Spatially Correlated Errors Can Mess Up The Fixed Effect You Love. *The American Statistician*, Volume 64, (4):325-334.

James, W. And Stein, C. (1960). Estimation with Quadratic Loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1), Berkeley, CA: *University Of California Press*, Pp. 361-379.

Jean-Paul Chiles and Pierre Delfiner (1999). Geostatistics: Modelling Spatial Uncertainty, 2nd Edition. ISBN: 978-0-470-18315-1.

Jeffreys, H. (1946). An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London. Series A*, Mathematical And Physical Sciences 186 (1007): 453–461.

John Mullahy. (1997). Heterogeneity, Excess Zeros, and the Structure of Count Data Models. *Journal of Applied Econometrics*, Vol. 12, 337-350.

John Snow. (1855). On the Mode of Communication of Cholera. *John Churchill*, New Burlington Street, England. *Journal Of The royal Statistical Society. Series D* (The Statistician), 34(3), Pp. 323-335.

Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of American Statistical Association*, Vol. 90, No. 430, 773-795.

Kass, R. and Steffy, D. (1989). Approximate Bayesian Inference In Conditionally Independent Hierarchical Models. *J. Am. Stat. Assoc*. 84, 717-726.

Keeling, M. J. (1999). The Effects of Local Spatial Structure on Epidemiological Invasions. *Proceedings of the Royal Society Biological Sciences Series B*, 266:859-867.

Kelsall, J. and Wakefield, J. (2002). Modelling Spatial Variation in Disease Risk: A Geostatistical Approach. *Journal of the American Statistical Association*. 97: 692–701.

Knorr-Held, L. and Raßer, G. (2000). Bayesian Detection of Clusters and Discontinuities in Disease Maps. *Biometrics*; 56: 13–21.

Knorr-Held, L. (2000). Bayesian Modelling Of Inseparable Space-Time Variation in Disease Risk. *Statistics in Medicine* 19, 2555-2567.

Knorr-Held, L. and Besag, J. (1998). Modelling Risk from a Disease in Time and Space. *Statistics in Medicine*; 17(18): 2045–2060.

Koopman, J.S. and Lynch J.W. (1999). Individual Causal Models and Population Systems Models in Epidemiology. *Am J Public Health* 89:1170–1174.

Kuan, P.F., Dongjun, C., Guangjin, P., James, A.T., Stewart, R. and Andkeleş, S. (2011). A Statistical Framework For The Analysis Of Chip-Seq Data. *J Am Stat Assoc*; 106(495):891–903.

Kulldoff, M. and Nagarwalla, N. (1995). Spatial Disease Clusters: Detection and Inference. *Statistics in Medicine*, 14(8):799-810.

Kulldorff, M., Athas, W., Feuer, E., Miller, B. and Key, C. (1998). Evaluating Cluster Alarms: A Space-Time Scan Statistic and Cluster Alarms in Los Alamos. *American Journal of Public Health*, 88: 1377-1380.

Kumar, V., Abbas, A. K., Fausto, N. and Mitchell, R. N. (2007). Robbins Basic Pathology (8th Ed.). *Saunders Elsevier*. Pp. 516–522. ISBN 978-1-4160-2973-1.

Laird, N. (1982). Empirical Bayes Estimates Using the Nonparametric Maximum Likelihood Estimate for the Prior. *J. Stat. Computer Simulation*, 15: 211-220.

Laird, N.M and Louis, T.A. (1989). Empirical Bayes Confidence Intervals Based On Bootstrap Samples. *Journal of American Statistical Association*, 82(25):739-750.

Lance, A. W., Bradley, P. C., Hong Xia, and Alan, E. G. (1997). Hierarchical Spatio-Temporal Mapping Of Disease Rates. *Journal of the American Statistical Association*, 92(438):607-617.

Last, J. M. (2001). Re: A Dictionary of Epidemiology, Fourth Edition, Edited By John M. Last, Robert A. Spasoff And Susan G. Harris. *American Journal of Epidemiology*; 154(4):389-392.

Lawn, S.D. And Zumla, A.I. (2011). Tuberculosis. *Lancet* 378 (9785): 57 –72.

Lawson, A, BöHning, D., Biggeri, A., Lesaffre, E. and, Viel J-F. (1999). Disease Mapping and Its Uses. In Lawson A, Biggeri A, BöHning D, Lesaffre E, Viel J-F, Bertollini R Eds. Disease Mapping And Risk Assesment For Public Health. Chichester, UK: *John Wiley & Sons*: 3–13.

Lawson, A. (2000). Statistical Methods in Spatial Epidemiology. *John Wiley & Sons*, Sussex.

Lee, D., Rushworth, A. and Sahu, S. (2014). A Bayesian Localised Conditional Auto-Regressive Model For Estimating The Health Effects Of Air Pollution. *Biometrics* 70, 419-429.

Lee, P. (1997). Bayesian Statistics: An Introduction. 2nd Ed. *London*: Arnold.

Leroux, B., Lei, X. and Breslow. N. (1999). Estimation Of Disease Rates In Small Areas: A New Mixed Model For Spatial Dependence, Chapter Statistical Models In Epidemiology, The Environment And Clinical Trials, Halloran, M and Berry, D (Eds), Pp. 135-178. *Springer-Verlag*, New York.

Lesage J. and R K. Pace. (2009): Introduction to Spatial Econometrics. Boca Raton: *Chapman and Hall/CRC*. Chapter 5, Pp. 123-154.

Lesage, J. (1999). The Theory and Practice of Spatial Econometrics. Webb Manuscript (Http://Www.Spatial-Econometrics.Com). Department Of Economics, University Of Toledo.

Link, B. and Phelan, J. (1995). Social Conditions as Fundamental Causes of Disease. *Journal of Health and Social Behavior, Extra Issue*, 80–94.

Longini, I. M., Koopman, J.S., Monto, A.S. and Fox, J.P. (1982). Estimating Household and Community Transmission Parameters for Influenza. *American Journal of Epidemiology* 115: 736-751.

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2005).Geographic Information Systems and Science. Second Edition. New York: *Wiley*.

Lord, D. (2006). Modelling Motor Vehicle Crashes Using Poisson-Gamma Models: Examining The Effects Of Low Sample Mean Values And Small Sample Size On The Estimation Of The Fixed Dispersion Parameter. *Accident Analysis & Prevention*, Vol. 38, No. 4, Pp. 751-766.

Lu, H., Hodges, J.S. and Carlin, B.P. (2007). Measuring the Complexity of Generalized Linear Hierarchical Models. *Canadian Journal of Statistics.* 35:69–87.

Lu, H., Reilly, C., Banerjee, S. and Carlin, B. (2007). Bayesian Areal Wombling via Adjacency Modelling. *Environmental and Ecological Statistics* 14, 433-452.

Maciel, E.L.N., Pan, W., Dietze, R., Peres, R.L., Vinhas, S.A., Ribeiro, F.K., Palaci, M., Rodrigues, R.R., Zandonade, E. and Golub,J.E.. (2010). Spatial Patterns Of Pulmonary Tuberculosis Incidence And Their Relationship To Socio-Economic Status In Vitoria, Brazil. *Int. J. Tuberc. Lung Dis*. 14(11): 1395–1402.

Macnab, Y., Farrell, P. Gustafson, and Wen, S. (2004). Estimation in Bayesian Disease Mapping. *Biometrics* 60: 865-873.

Macnab, Y.C. (2003). Hierarchical Bayesian Modelling Of Spatially Correlated Health Service Outcome and Utilization Rates. *Biometrics*, 59(2):305–16.

Macnab, Y.C. and Dean, C.B. (2000). Parametric Bootstrap and Penalized Quasi-Likelihood Inference in Conditional Autoregressive Models. *Statistics in Medicine*. 19:2421–2435.

Macnab, Y.C. and Dean, C.B. (2001). Autoregressive Spatial Smoothing and Temporal Spline Smoothing For Mapping Rates. *Biometrics*. 57: 949-956.

Manda, S.M, Feltbower, R.G. and Gilthorpe, M.S. (2011). Review and Empirical Comparison of Joint Mapping of Multiple Diseases. *Southern African Journal of Epidemiology and Infection,* 27(4):169-182.

156. Marshall, R.J. (1991). A Review of Methods for the Statistical Analysis of Spatial Patterns of Disease. *Journal of the Royal Statistical Society Series A* (Statistics in Society); 154: 421–41.

Martinez-Beneito, M.A., Lopez-Quilez, A. and Botella-Rocamora, P. (2008). An Autoregressive Approach to Spatio-Temporal Disease Mapping. *Statistics in Medicine*, 27 (15):2874-2889.

Mccullagh, P. and Nelder, J. A. (1989). Generalized Linear Models. *Chapman & Hall,* Second Edition.

Militino, A. F., Ugarte, M.D. and Garcia-Reinaldos, L. (2004). Alternative Models for Describing Spatial Dependence among Dwelling Selling Prices. *Journal of Real Estate Finance and Economics*, 29: 193–209.

Mohammadreza, M., Rory, W. and Andrew Forbes. (2014). Disease Mapping And Regression With Count Data In The Presence Of Over-dispersion And Spatial Autocorrelation: A Bayesian Model Averaging Approach. *Int. J. Environ. Res. Public Health*, 11: 883-902.

Mollie, A. (1996). Bayesian Mapping of Disease. In: Gilks, Walter R., Richardson, Sylvia, and Spiegelhalter, David J. (Editors), Markov Chain Monte Carlo in Practice. *New York: Chapman & Hall*. Pp 359–79.

Moran, P. A. P. (1950). Notes On Continuous Stochastic Phenomena. *Biometrika* 37(1): 17–23.

Musenge, E. (2013). Modelling Spatiotemporal Patterns Of Childhood HIV/TB Related Mortality And Malnutrition: Applications To Agincourt Data In Rural South Africa. A (Phd) Thesis Submitted To The Faculty Of Health Sciences, University Of The Witwatersrand, Biostatistics And Epidemiology.

Musenge, E. Vounatsou, P. and Kahn, K. (2011). Space-Time Confounding Adjusted Determinants Of Child HIV/TB Mortality For Large Zero-Inflated Data In Rural South Africa. *Spatial Spatio-Temporal Epidemiology*; 2(4): 205-17.

Musenge, E., Tobias, F.C., Kathleen, K. and Penelope, V. (2013). Bayesian Analysis Of Zero Inflated Spatiotemporal HIV/TB Child Mortality Data Through The INLA And SPDE Approaches: Applied To Data Observed Between 1992 And 2010 In Rural North East South Africa. *International Journal of Applied Earth Observation and Geoinformation*. Volume 22: 86–98.

National Department of Health. (2010). Management of Drug-Resistant Tuberculosis: Policy Guidelines. Http:// Www.Tbonline.Info/Media/Uploads/Documents/Mdr-Tb_Sa_2010.Pdf (Accessed 4 November 2013).

Nelder, J, A. and Wedderburn, W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, Series A* 135: 370-384.

Ngesa Oscar Owino (2014). Bayesian Spatial Models with Application to HIV, TB and STI Modelling in Kenya (A Phd Thesis). University Of Kwazulu-Natal, South Africa.

Noel Cressie. (1993). Statistics for Spatial Data, Revised Edition. ISBN: 978-0-471-0255-0.

Noriszura, I, L and Abdul, A. (2007). Handling Over-dispersion With Negative Binomial And Generalized Poisson Regression Models. Casualty Actuarial Society *Forum*, Winter 2007.

Ntzoufras, I. (2011). Bayesian Modelling Using Winbugs, Volume 698. *Wiley*.

Nunes, C. (2007). Tuberculosis Incidence in Portugal: Spatio-Temporal Clustering. *International Journal of Health Geography*, 6:30.

Onozuka, D. and Hagihara, A. (2007). Geographic Prediction of Tuberculosis Clusters in Fukuoka, Japan Using the Space-Time Scan Statistic. *BMC Infectious Diseases*, 7:26.

Osei Peprah. (2010). Statistical Methods for Disease Mapping. Submitted In Partial Fulfillment Of A Postgraduate Diploma, *African Institute For Mathematical Sciences* (AIMS).

Palm, T.A. (1890). The Geographical Distribution and Aetiology of Rickets. *Practitioner*. 45:270-279.

Palmgren, Juni. (1981). The Fisher Information Matrix For Log-Linear Models Arguing Conditionally In The Observed Explanatory Variables. *Biometrika* 68: 563-566.

Paul, D. D., Marin, P. and Dragos, B. (2013). Comparative study regarding the methods of interpolation. Recent Advances in Geodesy and Geomatics engineering. ISBN: 978-960-474-335-3.

Pascutto, C., Wakefield, J. C., Best, N. G., Richardson, S., Bernardinelli, L., Staines, A. and Elliott, P. (2000). Statistical Issues in the Analysis of Disease Mapping Data. *Statistics in Medicine*. 19: 2493–519.

Paul, C., Rhind, S.M., Kyle, C.E., Scott, H., Mckinnell, C. and Sharpe, R.M. (2005). Cellular And Hormonal Disruption Of Fetal Testis Development In Sheep Reared On Pasture Treated With Sewage Sludge. *Environmental Health Perspectives* 113: 1580–1587.

Pettitt, A.N., Weir, I.S. And Hart, A.G. (2002). A Conditional Autoregressive Gaussian Process For Irregularly Spaced Multivariate Data With Application To Modelling Large Sets Of Binary Data. *Statistics And Computing*, 12: 353–367.

Piegorsch, W. W. (1990). Maximum-Likelihood Estimation for the Negative Binomial Dispersion Parameter. *Biometrics* 46: 863–867.

Poch, M., and Mannering, F. L. (1996). Negative Binomial Analysis of Intersection Accident Frequency. *Journal of Transportation Engineering*, Vol. 122, No. 2, Pp.105-113.

Renato Assuncao and Elias Krainski. (2009). Neighborhood Dependence in Bayesian Spatial Models. *Biometrical Journal* 51, 5, 851–869 DOI: 10.1002/Bimj.200900056.

Rezaeian M., Dunn, G., St Leger, S. and Appleby, L. (2007). Geographical Epidemiology, Spatial Analysis and Geographical Information Systems: A Multidisciplinary Glossary. *Journal of Epidemiology of Community Health*. Volume 61(2):98–102. DOI: 10.1136/ Jech.2005.043117.

Richard Webster and Margaret A. Oliver. (2001). Geostatistics for Environmental Scientists, 2nd Edition. ISBN: 978-0-470-02858-2.

Richardson, S. (2003). Spatial Models in Epidemiological Applications. In Green PJ, Hjort NL, Richardson S Eds. Highly Structured Stochastic Systems. Oxford: *Oxford University Press:* 237–259.

Richardson, S., Guihenneuc, C. and Lasserre, V. (1992). Spatial Linear Models with Autocorrelated Error Structure. *The Statistician*, 41: 539–57.

Rindra, V.,Vincent, R., Fanjasoa, R., Philippe, S. and Dominique, J.B. (2010). Bayesian Mapping of Pulmonary Tuberculosis in Antananarivo, Madagascar. *BMC Infectious Diseases.*10:21.

Ripley, B. D. (1977). Modelling Spatial Patterns. *Journal of the Royal Statistical Society. Series B* (*Methodological*). Volume 39. Issue 2, 172-212.

Ripley, B.D. (1981). Spatial Statistics. *Wiley Sons, New York. ISBN*: 9780471083672. DOI: 10.1002/0471725218.

Robert, C. (1996). Mixtures of Distributions: Inferences and Estimation. In: Markov Chain Monte Carlo in Practice, Gilks, W., Richardson, S. and Spieglehalter, D. (Eds.), Boca Raton: *Chapman & Hall/CRC*.

Robinson, M. D. and Smyth, G. K.(2007). Small Sample Estimation of Negative Binomial Dispersion, With Applications to SAGE Data. *Biostatistics*. Vol. 9, Issue 2, Pp. 321-332.

Rodriguez, G. (2013). Models For Count Data With Over-Dispersion. Addendum to the WWS 509 Notes on Extra-Poisson Variation and the Negative Binomial Model.

Roger Levy and Klinton Bicknell. (2012). The Utility Of Modelling Word Identification From Visual Input Within Models Of Eye Movements In Reading. *Visual Cognition*, 20 (4–5), 422–456.

Ross, G.J.S., and Preece, D. A. (1985). The Negative Binomial Distribution. *Journal of the Royal Statistical Society. Series D (The Statistician).* Vol. 34, No. 3 (1985), pp. 323-335.

Rue, H. And Held, L. (2005). Gaussian Markov Random Fields; Theory and Applications, Vol. 104 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian Inference For Latent Gaussian Models Using Integrated Nested Laplace Approximations (With Discussion). *Journal Of The Royal Statistical Society, Series B* 71: 319-392.

Saha, K. and Paul, S. (2005). Bias-Corrected Maximum Likelihood Estimator Of The Negative Binomial Dispersion Parameter. *Biometrics* 61: 179–185.

Sampurna. K., Chamnein, C. and Apiradee, L. (2014). Spatial and Temporal Variations in TB Incidence, Nepal. *South East Asian J Trop Med Public Health*. Vol 45 No. 1.

Sandwell, D. T. (1987). Biharmonic Spline Interpolation of Geos-3 and Seasat Altimeter Data, *Geophys. Res. Lett.*, 14: 139-142.

Sartorius, B., Kahn, K., Vounatsou, P., Collinson, M. and Tollman, S. (2010). Young and Vulnerable: Spatial-Temporal Trends and Risk Factors for Infant Mortality in Rural South Africa (Agincourt), 1992-2007. *BMC Publ Health* 2010: 10: 645.

Schrodle, B., Held, L., Riebler, R. and Danuser, J. (2011). Using Integrated Nested Laplace Approximations For The Evaluation Of Veterinary Surveillance Data From Switzerland: A Case Study. *Journal of the Royal Statistical Society Series C*, 60: 261-279.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics.* 6: 461-464.

Shannon, W. D. (2008). Cluster Analysis. Handbook of Statistics, Vol. 27. ISSN: 0169-7161. *Published By Elsevier* B.V. DOI: 10.1016/S0169-7161(07)27011-7.

Shen, W, and Louis, T. A. (1998). Triple-Goal Estimates in Two-Stage, Hierarchical Models. *Journal of the Royal Statistical Society, Series B*. 60: 455–471. 916, 917, 918.

Small, P. M., and Pai, M. (2010). Tuberculosis Diagnosis—Time For A Game Change. N. *Engl. J. Med*. 363:1070–1071.

Smans, M. & Esteve, J. (1992). Practical approaches to disease mapping, in Geographical and Environmental Epidemiology: Methods for Small-Area Studies, P. Elliott, J. Cuzick, D. English & R. Stern, eds. Oxford University Press, Oxford, pp. 141–157.

Smith, T. Spiegelhalter, D. and Thomas, A. (1995). Bayesian Approaches to Random-Effects Meta-Analysis: A Comparative Study. *Statistics in Medicine,* Volume 14, 2685-2699.

Snijders, T.A.B. and Roel J. Bosker. (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling. *Sage Publications*. ISBN 0761958908, 9780761958901.

Souza, W.V., Marilia, S.C., Ximenes, R.A.A., Albuquerque, M.F.M.And Christovam, C.B. (2007). Tuberculosis in Intra-Urban Settings: A Bayesian Approach. *Tropical Medicine and International Health*. Volume 12 No 3: 323–330.

Souza, W.V., Ximenes, R.A.A., Albuquerque, M.F.M. et al. (2000). The Use Of Socioeconomic Factors In Mapping Tuberculosis Risk Areas In A City Of Northern Eastern Brazil. *Revista Panamericana De Salud Publica,* 8: 403-410.

Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. (1994, 2003). BUGS: Bayesian Inference Using Gibbs Sampling. *MRC Biostatistics Unit*, Cambridge.

Spiegelhalter, D., Thomas, A., Best, N and Lunn, D. (2007). Winbugs User Manual Version 1.4, January 2003. Upgraded To Version 1.4.3.

Spielgelhalter, D., Best, N. and Carlin, B. (1998). Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrarily Complex Models. *MRC Biostatistics Unit*, Cambridge.

Spyrou, C., Stark, R., Lynch A.G. and Tavare, S. (2009). Bayes Peak: Bayesian Analysis of Chip-Seqdata. *BMC Bioinformatics*, 10,299.

Srinivasan, R. and Venkatesan, P. (2013). Bayesian Kriging of Tuberculosis in Chennai: A Small Scale Analysis. *Indian Journal Of Applied Research*, Volume: 3 Issue: 7 ISSN - 2249-555X.

Statistics South Africa (2013). Mortality and Causes of Death in South Africa, 2010: Findings from Death Notification. (Www.Statssa.Gov.Za - See More At: Http://Www.Tbfacts.Org/Tb-Statistics-South-Africa).

Statistics South Africa. (2007). Mortality and Causes of Death in South Africa: Findings from Death Notification. Statistical Release PO309.3. Pretoria: Statistics South Africa, 2009. Available From: Http://Www.Statssa.Gov.Za/Publications/P03093/P030932007.Pdf.

Stead, W. W., Lofgren, J. P. and Senner, J. W. (1994). Invited Commentary: Relative Susceptibility Of Black Americans To Tuberculosis [Comment]. *American Journal of Epidemiology*. 139: 531±2; Discussion 533±4.

Stead, W. W., Senner, J. W., Reddick, W. T. and Lofgren, J.P. (1990). Racial Differences In Susceptibility To Infection By Mycobacterium Tuberculosis. *N. Engl. J. Med*., 322,422±427.

Stephen W. Raudenbush and Anthony S. Bryk (2002): Hierarchical Linear Models: Applications And Data Analysis Methods (Advanced Quantitative Techniques In The Social Sciences) 2nd Edition. *Sage Publications Ltd*. London RC2A 4PU.

Sun, D., Tsutakawa, R. K. and Speckman, P. L. (1999). Bayesian Inference For CAR (1) Models With Noninformative Priors. *Biometrika*; 86:341–50.

Sun, D., Tsutakawa, R. K. and Speckman, P. L. (1999). Posterior Distribution Of Hierarchical Models Using CAR (1) Distributions. *Biometrika*, 86: 341–350.

Sun, D., Tsutakawa, R. K., Kim, H. and He, Z. (2000). Spatio-Temporal Interaction with Disease Mapping. *Statistics in Medicine*: 19: 2015–35.

Tiwari, N., Kandpal, V., Tewari, A., Ram Mohan Rao, K. and Tolia, V.S. (2010). Investigation of Tuberculosis Clusters in Dehradun City of India. *Asian Pacific Journal of Tropical Medicine*, 486-490.

Ugarte, M, D., Goicoa, T. and Militino, F.A. (2010). Spatio-Temporal Modelling Of Mortality Risks Using Penalized Splines. *Environmetrics*, 21: 270-289.

Ugarte, M, D., Militino, F.A. and T. Goicoa, T. (2008). Prediction Error Estimators in Empirical Bayes Disease Mapping. *Environmetrics*, 19: 287-300.

Venkatesan, P. and Srinivasan, R. (2010).Modelling The Spatial Variogram of Tuberculosis for Chennai Ward in India. *Indian Journal of Science and Technology*. Vol. 3 No. 2. ISSN: 0974- 6846.

Venkatesan, P., Srinivasan, R., and Dharuman, C. (2010). Bayesian Conditional Autoregressive Model For Mapping Tuberculosis Prevalence In India. . *International Journal of Pharmaceutical Studies and Research* IJPSR/Vol. III/ Issue I/January-March, 2012/01-03. E-ISSN 2229-4619.

Venkatesan, P., Srinivasan, R., and Dharuman, C. (2012). Bayesian Conditional Autoregressive Model for Mapping Tuberculosis Prevalence in India. *International Journal of Pharmaceutical Studies and Research*. IJPSR.Vol. III. Issue I.E-ISSN 2229-4619.

Venkatesan,P. and Srinivasan, R. (2008). Bayesian Model of HIV/AIDS in India: A Spatial Analysis, Applied *Bayesian Statistical Analysis*, Page No. 51-56.

Verver, S., Warren, R.M., Munch, Z. (2004). Transmission of Tuberculosis in A High Incidence Urban Community In South Africa. *International Journal of Epidemiology*, 33: 351–357.

Victor De Oliveira. (2012). Bayesian Analysis of Conditional Autoregressive Models. *Annals of Institute of Statistical Math* (2012) 64:107–133.

Vieira, R.C., Prado, T.N., Siqueira, M.G., Dietze, R. and Maciel, E.L. (2008). Spatial Distribution of New Tuberculosis Cases in Vitória, State Of Espírito Santo, Between 2000 and 2005. *Rev Soc Bras Med Trop*. 41:82–86.

Wakefield, J. (2007). Disease Mapping and Spatial Regression with Count Data. *Biostatistics,* 8(2):158-183.

Wakefield, J. C., Best, N. G. and Waller, L. A. (2000). Bayesian Approaches to Disease Mapping. In: Elliott, P.,Wakefield, J. C., Best, N. G. and Briggs, D. (Editors), Spatial Epidemiology: Methods and Applications. Oxford: *Oxford University Press*. Pp 104–27.

Wall, M.M. (2002). A Close Look At The Spatial Structure Implied By The CAR And SAR Models. *Journal of Statistical Planning and Inference,* 121: (2004) 311 – 324.

Waller, L., B. Carlin, H. Xia, and A. Gelfand (1997). Hierarchical Spatiotemporal Mapping Of Disease Rates. *Journal of the American Statistical Association* 92: 607-617.

Waller, L.A. and Gotway, C.A. (2004). Applied Spatial Statistics For Public Health Data, *Wiley-Interscience*.Volume 368: ISBN: 978-0-471-38771-8.

Walter, S. D. (2000). Disease Mapping: A Historical Perspective. In: Elliott, P., Wakefield, J. C., Best, N. G. And Briggs, D. (Editors), Spatial Epidemiology: Methods and Applications. *Oxford University Press*. Pp 223–39.

Wasserman, L. (2003). All Of Statistics: A Concise Course In Statistical Inference. *New York: Springer*.

Wayner V. Souz, Marilia S. Carvalho, Maria De Fátima P. M. Albuquerque, Christovam C. Barcellos and Ricardo A. A. Ximenes. (2007). Tuberculosis in Intra-Urban Settings: A Bayesian Approach. *Tropical Medicine and International Health*. Volume 12; 323–330.

West, M. (1992). Modelling With Mixtures. Pp 503–524 In: Bernardo, J., Berger, J., David, A. And Smith, A. (Eds.). *Bayesian Statistics* 4. New York, NY: OUP.

Whittle. P. (1954). On Stationary Process in the Plane. *Biometrika*; 41 (3-4): 434-449.

Wikle, C.K., Berliner, L.M., and Cressie, N. (1998). Hierarchical Bayesian Space-Time Models. *Journal of Environmental and Ecological Statistics* 5:117–154.

Wikle, C.K., Milliff, R.F., Nychka, D. and Berliner, L.M. (2001). Spatio-Temporal Hierarchical Bayesian Modelling: Tropical Ocean Surface Winds. *Journal of the American Statistical Association* 96:382-39.

Williams, D. R. (1990). Socioeconomic Differentials in Health: A Review and Redirection. Social Psychology Quarterly, 53, 81–99.

Williams, D. R. (1997). Race and Health: Basic Questions, Emerging Directions. *Annals of Epidemiology*, 7; 322–333.

Willson, L. J., Folks, J. L., and Young, J. H. (1986). Complete Sufficiency and Maximum Likelihood Estimation for the Two-Parameter Negative Binomial Distribution. *Metrika* 33, 349-362.

Willson, L. J., Folks, J.L. and Young, J.H. (1984). Multistage Estimation Compared With Fixed-Sample-Size Estimation Of The Negative Binomial Parameter K. *Biometrics* 40: 109–117.

Wilson, M.E. (1995). Travel and the Emergence of Infectious Diseases. *Emerging Infectious Diseases*, 1:39-46.

Winkelmann, R. (2003). Econometric Analysis of Count Data. 4th Rev. Ed. Berlin: *Springer.*

World Health Organisation (WHO) Geneva, (2014). Global Tuberculosis Control 2014. (www.Who.Int/Tb/Publications/Global_Report/ - See More At: Http://Www.Tbfacts.Org/Tb-Statistics-South-Africa).

World Health Organization. (2002). Tuberculosis.

World Health Organization. (2010). Tuberculosis Fact Sheet N°104".

Xin-Xu Li, Li-Xia, W., Hui, Z., Shi-Wen, J., Qun, F., Jia-Xu, C. and Xiao-Nong, Z. (2014). Spatial Variations of Pulmonary Tuberculosis Prevalence Co-Impacted by Socio-Economic and Geographic Factors in People's Republic Of China, 2010. *BMC Public Health,* 14:257.

Xinyan, Z., Himel, M., Zaixiang, T., Lei, Z., Xiangqin, C., Andrew, K. B. and Nengjun, Y. (2017). Negative Binomial Mixed Models For Analyzing Microbiome Count Data. BMC Bioinformatics DOI 10.1186/S12859-016-1441-7.

Zaragoza Bastida A., Hernandez Tellez M., Bustamante Montes L.P., Medina Torres I., Jaramillo and Paniagua, J.N. Editors. (2012). Spatial and Temporal Distribution of Tuberculosis in the State Of Mexico. *Scientific World Journal*: 570278.

Zimmerman, D.L. (1993). Another Look at Anisotrophy in Geostatistics. *Mathematical Geology*, 25:453-470.

Appendix I.

R codes used for the multilevel models.

```
# Packages required
library("MASS")
library("lattice")
library("ctv")
library("sp")
library(maptools)
library(rgdal)
require(RColorBrewer)
library(spdep)
require(INLA)
###########Linearity#####################
############################
#########Model-1############
############################
formula<-TB~GINI+Poverty+UNEMP
result1<-
 inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE,mlik=TRUE,cpo=TRUE))
summary (result1)
```

```
############################
#########Model-2############
############################
formula<-TB~POPDEN+Noschool+AVGHouse
result2<-inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE),control.predictor
 = list(compute = TRUE))
summary(result2)
stru=result2$summary.random$District1
stru
############################
#########Model-3############
############################
formula<-TB~GINI+Poverty+UNEMP+POPDEN+Noschool+AVGHouse
result3<-inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE),control.predictor
 = list(compute = TRUE))
summary(result3)
unstru=result3$summary.random$District
unstru
############################
#########Model-4############
############################
formula<-TB~GINI+Poverty+UNEMP+f(District,model="iid")+f(District1,scale.model=TRUE,
 model="Besag",graph="dav.graph")
result4<-inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE),control.predictor
 = list(compute = TRUE))
summary(result4)
```

```
 stru=result4$summary.random$District1
stru
unstru=result4$summary.random$District
unstru
dav.graph$NUNSTR<-stru$"0.975quant"
spplot(dav.graph,"NUNSTR",   col.regions=bpy.colors(20))
unstru=result3$summary.random$District
unstru
dav.graph$UNUNSTR<-unstru$"0.975quant"
spplot(dav.graph,"UNUNSTR",   col.regions=bpy.colors(20))
############################
#########Model-5############
############################
formula<-TB~POPDEN+Noschool+AVGHouse+f(District,   model="iid")+f(District1,scale.model=TRUE,
 model="Besag",graph="dav.graph")
result5<-inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE),control.predictor
 = list(compute = TRUE))
summary(result5)
############################
#########Model-6############
############################
formula<-TB~GINI+Poverty+UNEMP+POPDEN+Noschool+AVGHouse+f(District,
 model="iid")+f(District1,scale.model=TRUE, model="Besag",graph="dav.graph")
result6<-
 inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE,mlik=TRUE,cpo=TRUE))
summary(result6)
```

```
dav.graph$NUNSTR<-stru$"0.975quant"
spplot(dav.graph,"NUNSTR",   col.regions=bpy.colors(20))
unstru=result1$summary.random$District
unstru
dav.graph$UNUNSTR<-unstru$"0.975quant"
spplot(dav.graph,"UNUNSTR",   col.regions=bpy.colors(20))
############################
#########Model-7############
############################
formula<-TB~f(District,model="iid")+f(District1,scale.model=TRUE,
  model="Besag",graph="dav.graph")
result7<-inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE),control.predictor
  = list(compute = TRUE))
summary(result7)
#####SPATIAL EFFECTS-Structured####
stru=result6$summary.random$District1
stru
#####SPATIAL EFFECTS-unstructured####
unstru=result6$summary.random$District
unstru
#################PLOTS#############
#############STRU##############
dav.graph$NUNSTR<-stru$"0.5quant"
spplot(dav.graph,"NUNSTR",   col.regions=bpy.colors(20))
dav.graph$NUNSTR1<-stru$"0.025quant"
spplot(dav.graph,"NUNSTR1",   col.regions=bpy.colors(20))
```

```
dav.graph$NUNSTR1<-stru$"0.975quant"
spplot(dav.graph,"NUNSTR1",   col.regions=bpy.colors(20))
##############UNSTRU#################
dav.graph$UNUNSTR<-unstru$"0.5quant"
spplot(dav.graph,"UNUNSTR",   col.regions=bpy.colors(20))
```

 # R-INLA codes for spatially weighted ICAR

```
############################
#########Model-5############
############################
formula<-TB~log(pop)+f(District,model="iid")+f(District1,scale.model=TRUE,
model="besag2",graph="dav.graph")
result5<-
inla(formula,family="poisson",data=dav_data,control.compute=list(dic=TRUE),control.predictor =
list(compute = TRUE))
summary(result5)
stru=result5$summary.random$District1
stru
unstru=result5$summary.random$District
unstru
dav.graph$NUNSTR<-stru$"0.975quant"
spplot(dav.graph,"NUNSTR",   col.regions=bpy.colors(20))
dav.graph$UNUNSTR<-unstru$"0.975quant"
spplot(dav.graph,"UNUNSTR",   col.regions=bpy.colors(20))
```

Appendix II

Table A1: TB prevalence data in the Eastern Cape Province from 2012 to 2015.

	Locations	LONG (X)	LAT (Y)	tb cases (Z):-	2012	2013	2014	2015	
1	Maluti	28.76	-30.31		791	784	843	875	
2	Umzimvubu	28.95	-30.77		2754	2169	2458	2931	
3	Amahlathi	27.26	-32.52		1514	1642	1579	1537	
4	Mbhashe	28.76	-32.16		994	1203	1185	1282	
5	Mnquma	28.01	-32.17		1115	1171	1272	1292	
6	Nkonkobe	26.83	-32.79		1139	1086	1042	1035	
7	Emalahleni	27.07	-31.55		855	744	777	713	
8	Engcobo	27.98	-31.67		1481	1321	926	770	
9	Intsika Yethu	27.64	-31.93		905	1026	737	833	
10	Inxuba Yethemba	25.37	-31.93		1092	1202	818	942	
11	Lukhanji	26.89	-32.1		1354	1328	1418	1222	
12	Sakhisizwe	27.83	-31.38		551	435	388	336	
13	Elundini	28.39	-30.91		547	505	611	475	
14	Maletswai	26.89	-30.92		457	489	533	539	
15	Sengu	27.64	-30.91		848	730	792	919	
16	Quakeni	29.69	-31.26		1682	1473	1355	1509	
17	King Dalindyebo	28.57	-31.7		2809	2892	2517	2154	
18	Mhlontlo	28.81	-31.19		1156	1028	1164	1056	
19	Nyandeni	29.13	-31.56		2847	2726	2838	2891	
20	Camdebbo	24.33	-32.23		973	905	899	1054	
21	Kouga	24.8	-33.91		2506	2706	2640	2529	
22	Makana	26.32	-33.24		1158	1135	1109	989	
23	A	27.64	-32.93		34	5409	5568	5147	
24	B	25.56	-33.74		3409	4016	3896	3862	

Table A2: TB 2014 Analysis Dataset.

District	GINI	Poverty	TB	Noschool	POPDEN	AVGHouse	UNEMP	Population	Logpop
1	0.69	0.8005	843	0.00094	61.68	3.7	20.69	268,428	5.428828
2	0.65	0.636	2458	0.0008	92.29	3.8	28.53	231,401	5.364365
3	0.71	0.5205	1579	0.001	28.46	3.5	30.86	121,432	5.084333
4	0.5	0.6843	1185	0.00212	88.68	4.1	37.11	270,865	5.432753
5	0.51	0.6405	1272	0.00114	94.34	3.5	26.46	312,077	5.494262
6	0.38	0.6923	1042	0.00072	37.16	3.4	49.24	138,513	5.141491
11	0.39	0.6253	777	0.00188	36.51	3.7	63.84	129,886	5.113562
12	0.72	0.6803	926	0.00197	62.78	4	41.78	141,911	5.152016
13	0.2	0.7201	737	0.00146	62.36	3.5	49.78	189,883	5.278486
14	0.5	0.3521	818	0.00107	4.39	3.4	25.68	50,924	4.706923
15	0.44	0.6195	1418	0.00078	51.91	3.5	36.97	221,281	5.344944
16	1.01	0.615	388	0.00127	25.78	3.7	52.99	57,997	4.763406
17	0.58	0.6363	611	0.00159	25.63	3.5	21.05	130,174	5.114524
18	0.64	0.1868	533	0.0011	10.42	3.4	26.99	45,463	4.657658
19	0.74	0.3944	792	0.00145	17.23	3.5	35.63	126,195	5.101042
23	0.48	0.6087	1355	0.00208	120.59	4.7	23.55	296,808	5.472476
21	0.62	0.4849	2517	0.0014	156.27	4	21.02	456,447	5.65939
22	0.64	0.599	1164	0.00147	87.33	4.2	24.05	246,656	5.392092
24	0.51	0.5907	2838	0.00182	133.71	4.6	40.61	330,461	5.51912
8	0.54	0.0931	899	0.0009	5.98	3.8	18.91	43,292	4.636408
9	0.72	0.2649	2640	0.00049	31.47	3.2	14.62	76,091	4.881333
10	0.54	0.4738	1109	0.00063	17.08	3.4	36.51	74,793	4.873861
7	0.54	0.4653	5568	0.00049	311.26	3.2	21.14	785,330	5.895052
20	0.69	0.442	3896	0.0003	569.32	3.4	26.77	1,111,767	6.046014

Appendix III

Ethical Clearance Certificate.

University of Fort Hare
Together in Excellence

ETHICAL CLEARANCE CERTIFICATE
REC-270710-028-RA Level 01

Certificate Reference Number:	QIN041SOBA01
Project title:	**Spatial modeling of Tuberculosis Epidemy in the demographics of Eastern Cape Province, South Africa**
Nature of Project:	PhD
Principal Researcher: Sub-Investigator:	Abiodun Davies Obaromi
Supervisor: Co-supervisor:	Prof Y Qin

On behalf of the University of Fort Hare's Research Ethics Committee (UREC) I hereby give ethical approval in respect of the undertakings contained in the above-mentioned project and research instrument(s). Should any other instruments be used, these require separate authorization. The Researcher may therefore commence with the research as from the date of this certificate, using the reference number indicated above.

Please note that the UREC must be informed immediately of

- Any material change in the conditions or undertakings mentioned in the document
- Any material breaches of ethical undertakings or events that impact upon the ethical conduct of the research

The Principal Researcher must report to the UREC in the prescribed format, where applicable, annually, and at the end of the project, in respect of ethical compliance.

Special conditions: Research that includes children as per the official regulations of the act must take the following into account:

Note: The UREC is aware of the provisions of s71 of the National Health Act 61 of 2003 and that matters pertaining to obtaining the Minister's consent are under discussion and remain unresolved. Nonetheless, as was decided at a meeting between the National Health Research Ethics Committee and stakeholders on 6 June 2013, university ethics committees may continue to grant ethical clearance for research involving children without the Minister's consent, provided that the prescripts of the previous rules have been met. This certificate is granted in terms of this agreement.

The UREC retains the right to

- Withdraw or amend this Ethical Clearance Certificate if
  - Any unethical principal or practices are revealed or suspected
  - Relevant information has been withheld or misrepresented
  - Regulatory changes of whatsoever nature so require
  - The conditions contained in the Certificate have not been adhered to

- Request access to any information or data at any time during the course or after completion of the project.

- In addition to the need to comply with the highest level of ethical conduct principle investigators must report back annually as an evaluation and monitoring mechanism on the progress being made by the research. Such a report must be sent to the Dean of Research's office

The Ethics Committee wished you well in your research.

Yours sincerely

Professor Gideon de Wet
Dean of Research

01 October 2015