# SPATIAL MODELLING OF TB PREVALENCE IN KENYA USING INTEGRATED NESTED LAPLACE APPROXIMATION (INLA)

BY

**BERNADETTE NJERI NZOMO**

**SCHOOL OF MATHEMATICS**
**COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES**
**UNIVERSITY OF NAIROBI**

**A Project Submitted in partial fulfillment for the requirement of the award of Master of Science in Biometry.**

**2014**

## DECLARATION

**CANDIDATE:**

I declare that this project is my original work and has not, wholly or in part, been submitted or presented for examination for an award of a degree in any other university.

Signature: _____Date: _____

Bernadette Njeri Nzomo  Reg.No I56/61294/2010

**SUPERVISOR:**

This project has been submitted for examination with my approval as University Supervisor

Signature: _____ Date: _____

Dr. Kipchirchir Isaac Chumba

School Of Mathematics

College Of Biological and Physical Sciences

University of Nairobi

## DEDICATION

This project is dedicated to: Lazarus Nzomo, Joyce Nzomo, Monicah Nzomo, Daniel Nzomo, Teresia Nzomo, Boniface Nzomo and Robinson Mwenda Njuki.

**ACKNOWLEDGEMENT**

First and foremost, I would like to give glory and honour to God for enabling me to conduct this research and write this project. My deepest gratitude goes to my supervisor Dr. Kipchirchir Isaac Chumba of the School of mathematics, University of Nairobi, for his scholarly inspiration, guidance, critic and support in conducting this research. Words can hardly describe my gratitude towards him. I also truly thank the Department of Statistics, School of mathematics University of Nairobi for offering me the opportunity to study in the institution.

I am also indebted to my confidant and a friend Idah Gacheri Micheni for the supportive encouragement and prayers during the research period. I am also grateful to my colleagues for their crucial support during the research period. I truly appreciate their input, advice and positive criticism during both the proposal and project writing. Many thanks also go to my closest friends Sylvia Njuki and Fridah Wangari Kamau for their advice and encouragement during the research period.

My heartfelt thanks go to the Kenya's National Division of Leprosy, TB and Lung Disease (DLTLD) expanded to National AIDS and STI control programme (NASCOP) for their cooperation during the research period. The successful completion of this study has indeed benefited from the input of many people in Ministry of Health (MOH) that I may not be able to mention. However, I acknowledge all their assistance. Thanks to all the people who contributed to this research in one way or another.

## ABSTRACT

Despite there being nationwide reduction in TB Cases incidence, there has been an increase in TB infection cases in various areas in the country  The use of registered numbers of reported new cases of TB is not enough to explain the distribution pattern and the determinants of the disease. Spatial statistics as method of analysis is a possible way to have the knowledge on the areas where the disease infection and distribution pattern is high.

In this study we apply Integrated Nested Laplace Approximation (INLA). INLA is a new tool for Bayesian inference on Latent Gaussian fields which substitute Markov Chain Monte Carlos (MCMC) simulations with deterministic approximations to the posterior marginals of interest. Its main benefits are the extreme speed and the high accuracy of the results.  Moreover INLA can be easily implemented through the INLA-program and its R-interface.

The study shows that TB prevalence is distributed different within the different provinces with Nyanza province having the largest number of the population with TB with a proportion of over 15% followed by Rift Valley province with a prevalence of between 10% to 15%, the province with the least number of people with TB is North Eastern province with a prevalence of less than 2%.  The study shows that there some areas in the country that have lesser health facilities, with North Eastern recording a negligible number of health facilities, this could imply that there may be more TB cases in the region, but due to lack of health facilities not all the active cases are registered

The main factors that affect the spread of TB from the study include; poverty, proportion taught on spread and control of TB, urban residence, all these factors are significant each lying in the 0.5 quant.  This implies that the government needs to put in place policy to improve the living standards of urban residents as well as the per capita income of her population.

## LIST OF ABBREVIATIONS AND ACCRONYMS

| | | |
|---|---|---|
| CAR | - | A conditional autoregressive |
| DLTLD | - | National Division of Leprosy, TB & Lung Disease |
| DOTS | - | Direct observation therapy |
| KLD | - | Kullback Leibler Deviance |
| MCMC | - | Markov Chain Monte Carlo Methods |
| MDR | - | Multi drug resistance |
| MOH | - | Ministry of Health |
| NASCOP | - | National AIDS and STI control programme |
| PET | - | Potential evapotranspiration |
| PPM | - | Public private Mix |
| SPDE | - | Stochastic Partial Differential Equation |
| TB | - | Tuberculoses |
| WHO | - | World Health Organization |

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

## 1.1 BACKGROUND INFORMATION

Disease maps have a long history; Snow (1854) used maps to determine the spread of cholera. Mapping of TB epidemic in Kenya will provide a rapid visual summary of complex geographic information and may identify subtle patterns in the data that are missed in tabular presentations. In this study, maps will be used variously for descriptive purpose, to generate hypotheses as to etiology, for surveillance to highlight areas at apparently high risk, and this could aid policy formation and resource allocation. Maps help place disease clusters and results of point-source studies in proper context (Wilkinson et al., 1997).

Disease maps typically show standardized morbidity ratios for geographical areas such as counties and districts. This can be calculated as the ratio of the observed positive cases or outcome of interest, in this study the observed positive case will be a TB reactive patient, to the expected number of number of positive cases (outcome) calculated by applying disease (TB) to the total population for the area of the study. Homogeneity within aggregate groups is important for meaningful interpretation. Openshaw (1984) described the modifiable area unit problem as, where there are different scales of measurements and different aggregation strategies that lead to different but equally valid maps that emphasize different features of the data. The main aim is to choose geographic units that are as small as possible and the choice will be dictated by the availability of data.

Bayesian statistics (Clayton and Kaldor, 1987) have been used to remove part of the random component from the map to give smoothed estimates of the relative risk in each area. Such estimates are a compromise between the local value of the standard morbidity ratio and the mean value for the map as a whole. Smoothing is greatest for the least-stable estimates. Although map smoothing on average produces a more stable and realistic map, an important issue is the extent to which disease (TB) excesses any truly high-risk areas for example those more sparsely populated might be smoothed away. The degree of smoothing will determine the trade-off

between high-risk areas correctly identified and areas without excess risk correctly identified (Richardson et al., 2004).

Most recently, Bayesian methods have become commonplace in epidemiology and the pharmaceutical industry, and they are becoming more widely accepted in Public Health practice. As early as 1993, review articles appeared extolling the virtues of MCMC in medical applications (Gilks et al., 1993). This increase in use has been facilitated by the implementation of software which provides a platform for the posterior distribution sampling which is necessary when relatively complex Bayesian models are employed. Basic ideas in Bayesian modeling stem from the extension of the likelihood paradigm to allow parameters within the likelihood model to have distributions. These distributions are called prior distributions. Thus parameters are allowed to be stochastic. By making this allowance, in turn, parameters in the prior distributions of the likelihood parameters can also be stochastic. Hence a natural parameter hierarchy is established. These hierarchical models form the basis of inference under the Bayesian paradigm. By combining the likelihood (data) model with suitable prior distributions for the parameters, a posterior distribution is formed which describes the behavior of the Bayesian Disease Mapping parameters after having seen the data.

The processes for mapping disease differ from random in the variation in the receptiveness of the study area to receive a point. In this study we shall apply approximate Bayesian inference using integrated nested Laplace approximations for latent Gaussian models. The map will provide an initial description of the geographic variation of TB risk in Kenya, and might help in choice and design of intervention, which is crucial for reducing the burden of TB in Kenya. Effective control requires evidence based utilization of resources. The type and degree of interventions need to be based on epidemiological patterns of TB risk. TB risk varies in space and time and hence it is important to describe the spatiotemporal variability of TB risk to guide control initiatives.

In the last decade maps have been used at different geographical scales in Sub Sahara Africa. In this study, analysis will seek to establish, to predict and map TB risk in Kenya using districts (old districts), as the referenced prevalence point. Existing risk maps are based on a theoretical

expert opinion or cultural and economic factors, (prior information) but these have limited information as they fail to provide insight into the transmission of TB in Kenya. It is important to characterize TB risk based on empirical evidence using TB specific indicator, in this case, TB prevalence infection in persons aged 15-49 years and assess its relationship with other risk factors. Prediction of risk based on point referenced data present some challenges when the data is sparsely distributed; such data exhibit autocorrelation, such that locations close to each other have similar risk. Hence models used should allow for spatial correlation, if otherwise the significance of the risk factors is overrated. Analysis of point reference data could be carried out using geostatistical models to attain optimal prediction. This approach allows simultaneous modeling of related issues such as risk assessment, spatial dependence, prediction and quantification of uncertainty. Accurate prediction can as well be introduced by including environmental factors likely to influence TB transmission.

In this study will apply the model-based geostatistical approach to analyze and predict TB risk in Kenya using referenced data from 71 districts in Kenya through the Ministry of Public Health and Sanitation, Division of Leprosy and TB. We shall adopt a Bayesian framework for inference and prediction, implemented using Integrated Nested Laplace Approximation INLA which replaces MCMC method. INLA is a new tool for Bayesian inference on Latent Gaussian fields which substitute MCMC simulations with deterministic approximations to the posterior marginals of interest. Its main benefits are the extreme speed and the high accuracy of the results. Moreover, INLA can be easily implemented through the INLA-program and its R-interface. Many models currently used in survival analysis can be cast into the Latent Gaussian fields framework and therefore be easily computed and analyzed.

Kenya ranks 13th on the list of 22 high-burden tuberculosis countries in the world and has the fifth highest burden in Africa. According to the World Health Organization's (WHO's) Global TB Report 2009, Kenya had approximately more than 132,000 new TB cases and an incidence rate of 142 new sputum smear-positive (SS+) cases per 100,000 population. Kenya's National Division of Leprosy, TB & Lung Disease (DLTLD) began to implement the WHO-recommended DOTS (Direct observation therapy) the internationally recommended strategy for

TB control, strategy in 1993 and reported 100 percent DOTS coverage by 1996. In 2005, the DOTS case detection rate reached WHO's target of 70 percent and rose to 72 percent in 2007. The DOTS treatment success rate also met WHO's target of 85 percent in 2007. Data from the national program show that Kenya had met the target for the treatment success rate in 2007. WHO estimates there were around 2,000 cases of multidrug-resistant (MDR) TB in Kenya in 2007, although only 4.1 percent of these cases were diagnosed and notified. There is a policy supporting MDR-TB diagnosis, treatment and a laboratory testing facility.

Kenya continues to treat more and more TB patients each year. However, widespread co-infection with HIV (close to 48 percent of new TB patients) makes TB treatment difficult. While the number of new cases appears to be declining, the number of patients requiring re-treatment has increased. The government placed the National Leprosy and Tuberculosis Program (NLTP) expanded to National Division of Leprosy, TB & Lung disease (DLTLD) and the national HIV/AIDS program in the same division in the Ministry of Health (MOH) to better address TBHIV/ AIDS co-infection. This resulted in increased collaborative TB-HIV/AIDS activities across the country. In 2007, the government demonstrated increased political commitment by upgrading the then-NLTP to a division within the MOH (DLTLD) and increased funding for TB control. With donor support, a greater proportion of TB patients benefited from improved DOTS services. The DLTLD implements TB-HIV/ AIDS treatment services, community-based DOTS (C-DOTS), and public-private mix (PPM) DOTS, as well as activities to address MDR-TB.

According to WHO/TB 1997, the re-emergence of infectious and communicable disease such as tuberculosis in Africa has been blamed on various problems such as poor epidemic surveillance; lack of medicine and medical services, weak health systems and poor health conditions. An estimated cumulative tuberculosis death stood at 30 million during the period 1990-1999 of which 6 million occurred in sub-Saharan Africa. Tuberculosis is perhaps the most important contagious disease in the World that causes millions of avoidable deaths worldwide.

## 1.1 STATEMENT OF THE PROBLEM

Despite there being nationwide reduction in TB Cases, there has been an increase in TB infection cases in various areas in the country. The use of registered numbers of reported new cases of TB is not enough to explain the distribution pattern and the determinants of the disease. Spatial statistics is a possible way to have the knowledge on the areas where the disease infection and distribution pattern has a high occurrence rate.

## 1.2 OBJECTIVES

The overall objective in this study is to identify the distribution patterns and determinants of TB prevalence in Kenya

The specific objectives that will be considered in the study are:

1. To examine the distribution pattern of TB using maps.
2. To examine the factors leading to spread and infection of TB.
3. To detect whether the set of locations observed contains clusters of events reflecting areas with associated increase in the likelihood of occurrence.

### 1.2.1 JUSTIFICATION OF THE STUDY

There will be additional knowledge added to the body of investigating disease prevalence. The method of data collection using geographic referencing (district are geo-referenced) is the best and most effective way of collecting TB data. There is need to identify the hotspots for TB prevalence in Kenya especially in most vulnerable locations for proper planning and administration.

# CHAPTER 2: LITERATURE REVIEW

Kazembe (2006), a model-based geostatistical methods were applied to analyze and predict malaria risk in areas where data was not observed. Topographical and climatic covariates were added in the model for risk assessment and improved prediction. A Bayesian approach was used for model fitting and prediction. The results; Bivariate models showed a significant association of malaria risk with elevation, annual maximum temperature, rainfall and potential evapotranspiration (PET). However in the prediction model, the spatial distribution of malaria risk was associated with elevation, and marginally with maximum temperature and PET. The resulting map broadly agreed with expert opinion about the variation of risk in the country, and further showed marked variation even at local level. High risk areas were in the low-lying lake shore regions, while low risk was along the highlands in the country. The map provided an initial description of the geographic variation of malaria risk in Malawi, and might help in the choice and design of interventions, which is crucial for reducing the burden of malaria in Malawi.

Kazembe et al., (2006), this paper used Pediatric ward register data from Zomba district, Malawi, between 2002 and 2003, as a case study. Two spatial models were developed. The first was a Poisson model applied to analyze hospitalization and minimum mortality rates, with age and sex as covariates. The second was a logistic model applied to individual level data to analyze case-fatality rate, adjusting for individual covariates. Rates of malaria hospitalization and in-hospital mortality decreased with age. Case fatality rate was associated with distance, age, wet season and increased if the patient was referred to the hospital. Furthermore, death rate was high on first day, followed by relatively low rate as length of hospital stay increased. Both outcomes showed substantial spatial heterogeneity, which may be attributed to the varying determinants of malaria risk, health services availability and accessibility, and health seeking behaviour. The increased risk of mortality of children referred from primary health facilities may imply inadequate care being available at the referring facility or the referring facility is referring the more severe cases which are expected to have a higher case fatality rate. Improved prognosis as the length of hospital stay increased, suggest that appropriate care when available can save lives.

Reducing malaria burden may require integrated strategies encompassing availability of adequate care at primary facilities, introducing home or community case management as well as encouraging early referral, and reinforcing interventions to interrupt malaria transmission.

Marshall (1990), mapping disease and mortality rates using empirical Bayes estimators. A new empirical estimator with parameters simplify estimated by moments is compared by iterative alternatives suggested by Clayton and Kaldor (1987). These methods are shrinkage estimators, in which the crude disease rate is shrunk towards an overall regional rate, and are in this sense global and invariant to spatial configuration. However it seems unjustifiable in effect, to ignore the spatial aspect of the problem. A local shrinkage estimator is therefore also suggested in which the crude rate is shrunk towards a local, neighborhood rate. Comparison of the estimators is done by some simulation experiments and an example showing infant mortality in Auckland, New Zealand is presented. When disease is relatively rare a global estimator gives the lowest total mean-square error, but for diseases that are more common and where the underlying spatial pattern is not uniform the local estimator performs best.

Ying (2003), the study uses hierarchical spatial models for the analysis of the geographical distribution of a non-rare disease. The work is motivated by the need for ascertaining regional variations in health services outcomes and resource use and for assessing the potential sources of these variations. The models discussed in this paper readily accommodate random spatial effects and covariate effects. Bayesian inferential framework and implementation of a hybrid Markov chain Monte Carlo method for full Bayesian model inference is discussed. This paper presented statistical methods for small area mapping of non-rare disease rates whose binomial sampling variation could not be approximated by a Poisson model. It discussed Bayesian spatial modeling and spatial smoothing of relative odds ratios where local information relevant to the rate odds for each individual area and 'global' information relevant to the overall dispersion of the underlying spatial disease rates are integrated via a Markov random field Gaussian prior. Specifically, the hyperparameter characterizes extra-binomial variation while depicts the degree of spatial autocorrelation within the data, that is, the inclination for areas of immediate neighbors to have similar disease rates. The model framework readily accommodates covariate effects, which

makes it possible to quantify the influence of geographically structured risk factors on disease occurrence and to further investigate residual variation in small-area disease risks that likely result from latent or unexplained covariate effects. To enable meaningful and reliable comparison of spatial disease rate odds, the paper discussed full Bayesian inference. The full Bayesian methods enabled the researcher to thoroughly investigate various properties of the posterior distribution of any model parameter and to adequately assess the uncertainty associated with risk prediction. Methodological innovations in Markov chain Monte Carlo algorithms have enabled Bayesian statisticians to select, among various Bayesian computational tools, suitable algorithms for their particular analytic or application purposes.

Cressie and Chan (1989), spatial modeling of regional variables; accumulated sudden infant death syndrome data, from 1974-1978 and 1979-1984 for counties of North Carolina. After a spatial exploratory data analysis, Markov random –field model was fitted to the data. The spatial trend was to capture the large scale variation in the data, and the variance and spatial dependence were to capture the small-scale variation. The principal feature of these data was that they come as counts from known or estimated base. It can be shown that probabilities can numerically represent a set of rational beliefs, that there is a relationship between probability and information, and that Bayes rule provides a rational method for updating beliefs in light of new information.
The flexible modeling of disease could require switching between a variety of relatively complex models. In this case it is convenient to have an efficient and flexible posterior sampling method which could be applied across a variety of models. The efficient algorithms for this purpose were developed within the fields of physics and image processing to handle large scale problems in estimation.

Owino et al., (2008), the study used Durbin modeling to answer the listed questions: where are the poor and what are their characteristics. Petrucii et al., (2004), mapping poverty in Ecuador, the study applies autologistic regression model adjusting for literacy, household size, mortality rates, temperature, land among other covariates. The results were used for poverty alleviation programs, emergency response and food aid.

Integrated nested Laplace approximation (INLA) provides a fast and yet quite exact approach to fitting latent Gaussian models which comprise many statistical models, including models with temporal or spatial dependence structures (Rue et al.,2009). As a result, many complex models that previously required the use of time-consuming Markov chain Monte Carlo (MCMC) calculations can be fitted fast and conveniently. Log Gaussian Cox processes, a particularly flexible class of spatial point process models are a special case of latent Gaussian models. Rue et al., (2009), Illian et al.,(2012) and Illian and Rue (2010), showed that complex point process models, including hierarchically marked point processes may conveniently be fitted with INLA. Standard approaches to parameter estimation for complex models based on MCMC, for example, would be very cumbersome and computationally prohibitive. Fitting spatial point process models to some spatial patterns is computationally intensive due to – amongst other things – the large number of individual points in the data set (Burslem et al., (2001); Waagepetersen (2007); Waagepetersen and Guan (2011); Law, Illian et al.,( 2009)). In some applications difficulties arise since point patterns with only a very small number of points can be collected, due to logistic limitations (e.g. for reasons of accessibility). These patterns are sometimes too small to justify the modeling of a single pattern. However, if replicates exist, a joint model of all replicates with a factor that accounts for variability among replicates caused by different conditions on different days may be more suitable. Mixed effect models for replicated point patterns have recently been considered in a frequentist approach for Gibbs processes (Illian and Hendrichsen (2010)). In that approach, parameter estimation was based on the pseudolikelihood of a Gibbs process as well as maximum quasi-likelihood optimization.

The aim of this study is to use the geostatistical method to a situation in which the stochastic variation is known to be latent Gaussian. In current geostatistical practice, the most widely implemented methodology for coping with latent Gaussian problems is trans-Gaussian kriging (Cressie, 1991), which consists of applying standard Gaussian methods after marginal non-linear transformation. In contrast the proposal is to embed linear kriging methodology within a more general distributional framework, analogous to the embedding of the Gaussian linear, model for mutually independent data within the framework of the generalized linear model (Nelder and Wedderburn (1972); McCullagh and Nelder (1989)).

# CHAPTER 3 : METHODOLOGY

## 3.1 DATA DESCRIPTION

Data used in this study were obtained from the register of public, private and mission health facilities in Kenya between 2009 and 2010. There are over 5,000 health facilities in the region and these facilities are managed by the Ministry of Public Health and Sanitation, Private Investor and Mission Organization (The Kenya National Health facility database, 2008). Only the geo-reference health facilities will be mapped.

For this study, cases with primary diagnosis as TB positive from the hospital register were used. The register included patient age in years, area of residence; number taught about TB control and infection, Tested for HIV, poverty index will be used from the KNBS database. Based on the location of the facility in the province, each case was matched to one of the districts in the province. Only geo-referenced cases were included in the spatial analysis, using districts as the spatial unit. The population at risk for the age 15-24 years and 25-49 year and, residential district, projected from 2008 census were obtained from KNBS office. Data used in the study is hospital routine data collected by the Ministry of Health and Sanitation Division of Leprosy and TB through different Government health facilities across the nation. The data contain both qualitative and quantitative variables.

## 3.2 MODELS

### 3.2.1 BAYESIAN INFERENCE

The Bayesian methods provide, parameter estimates with good statistical properties of the observed data, prediction for missing data and forecast of future data. It also provides a computational framework for model estimation, selection and validation. Thus, this method goes beyond the formal task of induction for which the method is derived. The numeric values of population characteristics are typically expressed in terms of a parameter $\theta$ and a numeric description of the subset make up a data set $\underline{x}$. We note that before the dataset is obtained, the numeric values of both the population characteristics and the data set are uncertain. After a data set $\underline{x}$ is obtained the information it contains can reduce our uncertainty in the purpose of Bayesian inference. The sample space $S$ is the set of all possible datasets, from which a single

data set $\underline{x}$ will result. The parameter space $\Omega$ is the set of possible parameter values, from which we hope to identify the value that best represents the true population characteristics. The Bayes' rule does not tell us what our beliefs should be, but rather tells us how they should change after realizing new information.

The assumption of spatial correlation at a hierarchical level above the likelihood is a fundamental assumption often made in Bayesian small health data modeling. This means that the correlation appears in prior distributions rather than in the likelihood itself. Often parameters are given such priors and the interpretation of prior distribution is that they provide additional 'data' for a problem and hence they can be used to improve estimation of parameters. A prior distribution is improper if its normalizing constant is infinite, while impropriety is a limitation of any prior distribution, it is not necessarily the case that an improper prior will lead to impropriety in the posterior distribution. The posterior distribution can often be proper even with an improper prior specification. Prior distribution and likelihood provide two sources of information about a research problem. The likelihood informs about the parameter through the data, while the prior distributions inform via prior assumptions. For a large sample size, the likelihood will contribute more to the relative risk estimation. The product of the likelihood and the prior distributions is called the posterior distribution. This distribution describes the behavior of the parameters after the data are observed and prior assumptions are made. For statistical hypothesis testing, the use of simulation approach is most suitable because of the complexity inherent in spatial processes; it is sometimes difficult to derive a legitimate test statistic whose probability distribution is known. An alternative approach is to use the computer to simulate multiple random spatial patterns, the spatial statistic is calculated for each, and then displayed as a frequency distribution. This simulated sampling distribution can then be used to assess the probability of obtaining our observed value for the index if the pattern had been random.

**Likelihood function**

Likelihood for data $x_1, x_2, \ldots x_n$

$$L(\underline{x} \mid \underline{\theta}) = \prod_{i=1}^{n} f(x_i \mid \underline{\theta}) \tag{3.1}$$

Where $\underline{\theta}$ is a $k$-dimensional vector $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ and $f(x_i \mid \underline{\theta})$ is a probability density functions. It is assumed that the sample values $\underline{x}$ given the parameters are independent, and hence we take the product of individual contribution in (3.1). Thus, the data is assumed to be conditionally independent. It is important to note that in many spatial applications the data would not be unconditionally independent and would be in fact correlated. This is important in disease mapping application which will be employed in this study. The logarithm of the likelihood is

$$\log L(\underline{x} \mid \underline{\theta}) = \sum_{i=1}^{n} \log f(x_i \mid \underline{\theta}) \tag{3.2}$$

and this is useful in model development.

**Posterior distribution**

The product of the likelihood and the prior distribution is called the posterior distribution. This distribution describes the behaviour of the parameters after the data are observed and prior assumptions are made.

The posterior is defined as

$$g(\underline{\theta} \mid \underline{x}) = L(\underline{x} \mid \underline{\theta}) g(\underline{\theta}) / C \tag{3.3}$$

where

$$C = \int_{\Omega} L(\underline{x} \mid \underline{\theta}) g(\underline{\theta}) d\underline{\theta} \tag{3.4}$$

where $g(\theta)$ is the prior distribution of $\underline{\theta}$ . Equation (3.3) can be equivalently represented as

$$g(\underline{\theta} \mid \underline{x}) \propto L(\underline{x} \mid \underline{\theta}) g(\underline{\theta}). \tag{3.5}$$

In this study our model representation reduces to

$$g(\theta \mid \underline{x}) \propto L(\underline{x} \mid \theta) g(\theta) \tag{3.6}$$

where $g(\theta)$ is a gamma distribution with parameters $\alpha, \beta$, that is

$$g(\theta) = \begin{cases} \frac{\beta^{\alpha}}{\sqrt{\alpha}} \theta^{\alpha-1} \exp(-\theta\beta), & \theta > 0; \alpha, \beta > 0 \\ 0 & \text{, otherwise} \end{cases} \tag{3.7}$$

and we write

$$\theta \sim G(\alpha, \beta).$$

Moreover, $x_i \mid \theta \sim \text{Poisson}(\theta e_i)$ where $e_i$ is the relative risk so that

$$L(\underline{x} \mid \theta) = \prod_{i=1}^{n} (\theta e_i)^{x_i} \exp(-\theta e_i) / x_i! \tag{3.8}$$

and the posterior distribution for fixed $\alpha, \beta$ is

$$g(\theta \mid \underline{x}) \propto \theta^{\sum_{i=1}^{n} x_i + \alpha - 1} \exp(-\theta(\sum_{i=1}^{n} e_i + \beta)) \tag{3.9}$$

which is a gamma distribution with parameters

$$\sum_{i=1}^{n} x_i + \alpha \text{ and } \sum_{i=1}^{n} e_i + \beta .$$

The posterior and the prior distributions belong to the same family and hence the Gamma family is said to be conjugate to the Poisson.

In Bayesian inference if there is more than one level, then in the second level the prior's are referred to as hyper-priors and the corresponding parameters are called hyper-parameters. Illustration for a Poisson distribution is as follows: For the first level

$$x_i \mid \theta \sim \text{Poisson}(\theta e_i)$$

$$\theta \mid \alpha, \beta \sim G(\alpha, \beta)$$

and for the second level we have

$$\alpha \mid \upsilon \sim h_\alpha(\upsilon)$$

$$\beta \mid \omega \sim h_\beta(\omega)$$

### 3.2.2 MARKOV CHAIN MONTE CARLO (MCMC) METHODS

Models in disease mapping have two or more levels and the resulting complexity of the posterior distribution of the parameters requires the use of sampling algorithms. In addition, the flexible modeling of the disease (TB) could require switching between a variety of relatively complex models. In this case, it is convenient to have an efficient and flexible posterior sampling method which could be applied across a variety of models.

MCMC methods are a set of methods which use iterative simulation of parameter values within a Markov chain. The convergence of this chain to a stationary distribution, which is assumed to be the posterior distribution, must be assessed.

A major limitation towards more widespread implementation of Bayesian approaches is that obtaining the posterior distribution often requires the integration of high-dimensional functions.

This can be computationally very difficult, but several approaches short of direct integration have been proposed (Smith (1991); Evans and Swartz (1995); Tanner (1996)). We focus here on MCMC methods, which attempt to simulate direct draws from some complex distribution of interest. MCMC approaches are so-named because one uses the previous sample values to randomly generate the next sample value, generating a Markov chain (the transition probabilities between sample values are only a function of the most recent sample value). The realization in the early 1990's (Gelfand and Smith (1990)) that one particular MCMC method, the Gibbs sampler, is very widely applicable to a broad class of Bayesian problems has sparked a major increase in the application of Bayesian analysis, and this interest is likely to continue expanding for some time to come. MCMC methods have their roots in the Metropolis algorithm (Metropolis and Ulam (1949); Metropolis et al., (1953)), an attempt by physicists to compute complex integrals by expressing them as expectations for some distribution and then estimate this expectation by drawing samples from that distribution. The Gibbs sampler (Geman and Geman (1984)) has its origins in image processing. Excellent and detailed treatments of MCMC methods are found in Tanner (1996) and Draper (2000).

**Monte Carlo Integration**

The origin of Monte Carlo approach was a method developed by physicist to use random numbers generated to compute integrals. Suppose we wish to compute a complex integral

$$\int_a^b h(x)dx \tag{3.10}$$

If we can decompose $h(x)$ into the production function $p(x)$ and a probability density function $f(x)$ defined over the interval $(a,b)$, then we note that

$$\int_a^b p(x)f(x)dx = E\left[p(X)\right] \tag{3.11}$$

so that the integral can be expressed as an expectation of $p(X)$ over the density $f(x)$. Thus, if we draw a large number $X_1,...,X_n$ of random variables from the density $f(x)$, then equation (3.11) can be approximated by

$$E[p(X)] \approx \frac{1}{n}\sum_{i=1}^{n} p(x_i). \tag{3.12}$$

This is referred to as Monte Carlo integration.

Monte Carlo integration can be used to approximate posterior of marginal posterior distributions required for a Bayesian analysis. Consider the integral

$$I(y) = \int_a^b p(y \mid x)\, f(x) dx$$

$$= E[p(y \mid X)]$$

$$\approx \frac{1}{n}\sum_{i=1}^{n} p(y \mid x_i) = \hat{I}(y) \tag{3.13}$$

where $\hat{I}(y)$ is the estimate of $I(y)$ and $x_i$ are draws from the density $f(x)$. The estimated Monte Carlo standard error is given by

$$SE\left(\hat{I}(y)\right) = \sqrt{\sum_{i=1}^{n} (p(y \mid x_i) - \hat{I}(y))^2}. \tag{3.14}$$

**The Markov chain**

Let $X_t$ denote the value of a random variable at time $t$, and let the state space refer to the range of possible values of $X$. The random variable is a Markov process if the transition probabilities between different values in the state space depend only on the random variable's current state. Let $X_1, X_2, ..., X_n$ be random variables, we say that $X_t$ satisfies Markov condition if

$$P(X_{t+1} = s_j \mid X_0 = s_k, ..., X_t = s_i) = P(X_{t+1} = s_j \mid X_t = s_i). \tag{3.15}$$

Thus, for a Markov random variable the only information about the past needed to predict the future is current state of the random variable, knowledge of the values of earlier state do not change the transition probability. A Markov chain refers to a sequence of random variables $(X_1, ..., X_n)$ generated by a Markov chain. A particular chain is defined most critically by its transition probabilities,

$$P_{ij} = P(X_{t+1} = s_j \mid X_t = s_i) \tag{3.16}$$

which is the probability that a process $\{X_t\}$ at state $s_i$ at time $t$ moves to state $s_j$ at time $t+1$.

Let

$$\boldsymbol{\pi}_j(t) = P(X_t = s_j) \tag{3.17}$$

denote the probability that the chain is in state $j$ at time $t$, and let $\boldsymbol{\pi}(t)$ denote the row vector of the state space probabilities at time $t$. We start the chain by specifying a starting vector $\boldsymbol{\pi}(0)$. Often all elements of $\boldsymbol{\pi}(0)$ are zero except for a single element of 1, corresponding to the process starting in that particular state. As the chain progresses, the probability values get spread out over the possible state space.

The probability that the chain is in state $s_i$, at time $t+1$ is given by the Chapman-Kolmogorov equation, which sums over the probabilities of being in particular state at the current step and the transition probability from that state into state $s_i$,

$$\begin{aligned}\pi_i(t+1) &= P(X_{t+1} = s_i) \\ &= \sum_k P(X_{t+1} = s_i \mid X_t = s_k) P(X_t = s_k) \\ &= \sum_k p_{ki} \pi_k(t)\end{aligned} \tag{3.18}$$

Successive iteration of the Chapman- Kolmogorov equation describes the evolution of the chain, we can more compactly write the Chapman- Kolmogorov equation in matrix form as follows. Define the probability transition matrix P as the matrix whose $(i, j)th$ element is $p_{ij}$. The Chapman- Kolmogorov equation becomes

$$\boldsymbol{\pi}(t+1) = \boldsymbol{\pi}(t)P \tag{3.19}$$

Using the matrix form, we immediately see how to quickly iterate the Chapman- Kolmogorov equation, as

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(t-1)P = (\boldsymbol{\pi}(t-2)P)P = \boldsymbol{\pi}(t-2)P^2. \tag{3.20}$$

Continuing in this fashion shows that

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)P^t. \tag{3.21}$$

Defining the $n-$step transition probability $p_{ij}^{(n)}$ as the probability that the process is in the $j$ given that it started in state $i$ $n$ steps ago, that is

$$p_{ij}^{(n)} = P(X_{t+n} = s_j \mid X_t = s_i) \tag{3.22}$$

it immediately follows that $p_{ij}^{(n)}$ is just the $ij-th$ element of $P^n$.

A Markov chain is said to be irreducible if there exists a positive integer $n_{ij}$ such that $p_{ij}^{(n)} > 0$ for all $i, j$, that is, all states communicate with each other, as one can always go from any state.

### 3.2.3 THE GIBBS SAMPLER

The Gibbs sampler is a special case of Metropolis-Hastings sampling. The task remains to specify how to construct a Markov chain whose values converge to the target distribution. Key to the Gibbs sampler is that one only considers univariate conditional distribution – the distribution when all of the random variables but one are assigned fixed values. Such conditional distributions are far easier to simulate than complex joint distributions and usually have simpler forms (often being normal, or other common prior distributions). Thus, one simulates $n$ random variables sequentially from the $n$ univariate conditionals rather than generating a single $n$ dimensional vector in a single pass using the full joint distribution.

To introduce the Gibbs sampler, consider a bivariate random variable $(x, y)$, and suppose we wish to compute one or both marginals,

$$f(x) \text{ and } f(y).$$

The idea behind the sampler is that it is far easier to consider a sequence of conditional distributions,

$$f(x \mid y) \text{ and } f(y \mid x),$$

than it is to obtain the marginal by integration of the joint density $f(x, y)$, for example

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy. \tag{3.23}$$

The sampler starts with some initial value $y_0$ for $y$ and obtains $x_0$ by generating a random variable from the conditional distribution

$$f(x \mid y = y_0) \tag{3.24}$$

The sampler then uses $x_0$ to generate a new value of $y_1$, drawing from the conditional distribution based on the value $x_0$,

$$f(y \mid x = x_0). \tag{3.25}$$

In general, the sampler proceeds by drawing $x_t$ from

$$x_t \sim f(x \mid y = y_{t-1}) \tag{3.26}$$

and drawing $y_t$ from

$$y_t \sim f(y \mid x = x_t). \tag{3.27}$$

Repeating this process $K$ times, generates a Gibbs sequence of length $K$, where a subset of points $(x_j, y_j)$ for $1 \le j < K$ are taken as the simulated draws from the full joint distribution, one samples the chain after a sufficient burn-in to remove the effects of initial sampling values at set time points following the burn-in. The Gibbs sequence converges to a stationary distribution that is independent of the starting values, and this stationary distribution is the target distribution we are trying to simulate (Tierney, 1994). The Gibbs sampler can be easily fitted using R software. Below is a theoretical illustration.

Take $x^{(0)} = (x_1^{(0)}, ..., x_K^{(0)})$ from $f^{(0)}(x)$ with $f(x^{(0)}) > 0$, and iterate for $t = 1,2,...$

$$\begin{aligned}
&1.\ \text{Generate } x_1^{(t)} \sim f_1(x_1 \mid x_2^{(t-1)}, ..., x_K^{(t-1)}).\\
&\vdots\\
&k.\ \text{Generate } x_k^{(t)} \sim f_k(x_k \mid x_1^{(t)}, ..., x_{k-1}^{(t)}, x_{k+1}^{(t-1)}, ..., x_K^{(t-1)}).\\
&\vdots\\
&K.\ \text{Generate } x_K^{(t)} \sim f_K(x_K \mid x_1^{(t)}, ..., x_{K-1}^{(t)}).
\end{aligned} \tag{3.28}$$

Under mild regularity conditions, the distribution of $x^t = (x_1^t, ..., x_K^{(t)})'$ denoted by $f^{(t)}(x)$ will converge to $f(x)$.

### 3.2.4 MODEL BASED GEOSTATISTICAL APPROACH

Geostatistical methodology is applied to solve the problem of predicting the realized value of a linear function of a Gaussian spatial stochastic process based on observations $Y_i = S(x_i) + Z_i$ at sampling locations $x_i$, where the $Z_i$ are mutually independent , zero -mean Gaussian random variables .

The theoretical framework for our extension of geostatistical methods is that, conditionally on the unobserved process $S(X)$ observations at sample locations $x_i$ form a generalized linear model with the corresponding values $S(x_i)$ appearing as an offset effect in the linear predictor. We use a Bayesian inferential framework implementing via the INLA (which replaces Markov chain Monte Carlo method), to solve the prediction problem for non linear functions, making a proper allowance for the uncertainty in the estimation of any model parameters.

Kriging refers to a widely used method for interpolation or smoothing spatial data. Given a set of data $y_i, i = 1,.., n,$ at spatial locations the kriging predictor for the underlying spatial surface, $S(X)$ say takes the form

$$\hat{S}(X) = \sum_{i=1}^{n} w_i(X) y_i \tag{3.29}$$

Where the kriging weighs $w_i(X)$ are derived from the estimated mean and covariance structure of the data. For a model based derivation, we can assume that the data are generated by the model

$$Y_i = \mu + S(X_i) + Z_i \quad , \quad i = 1,..., n, \tag{3.30}$$

where $\mu$ is a constant mean effect, $S(X_i)$ is a stationary Gaussian process with $\mathrm{E}[S(X)] = 0$ and $Cov(S(X), S(X')) = \delta^2 \rho(X - X')$, where $\delta^2$ and $\rho$ are variance and correlation function respectively and $Z_i$ are mutually independent. An equivalent formulation is that, conditionally on $S(X_i)$, the $Y_i$ are mutually independent and

$$Y_i \mid S(X_i) \sim N(\mu + S(X_i), \tau^2). \tag{3.31}$$

These distributional assumptions are often not made explicit. However the linear predictor (3.29) might be regarded as a natural choice under Gaussian assumptions, since it then minimizes

$$E[(\hat{S}(X) - S(X))^2] \tag{3.32}$$

The bald statement that kriging is linear prediction conceals a large body of methodology, collectively known as geostatistics in acknowledgment of its origin in mineral exploration. Much of the early development of geostatistical methodology was undertaken by Matheron et al., (1970). More recent text book accounts include Journel and Huijbregs (1978) and Isaaks and Srivastava (1989). Parallel independent development in stochastic process prediction (Whittle, 1963) and in the analysis of spatial variation (Matern, 1960) eventually led to placing of geostatistical methods within the wider setting of spatial statistics. This is further explained in Cressie (1991) and Ripley (1981).

The aim of this study is to use the geostatistical method to a situation in which the stochastic variation is known to be latent Gaussian. In current geostatistical practice, the most widely implemented methodology for coping with latent Gaussian problems is trans-Gaussian kriging (Cressie, 1991), which consists of applying standard Gaussian methods after marginal non-linear transformation. In contrast, the proposal is to embed linear kriging methodology within a more general distributional framework, analogous to the embedding of the Gaussian linear, model for mutually independent data within the framework of the generalized linear model (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989).

**Spatial Correlation**

Within spatial application it is often found that correlation will exist between spatial units. This correlation is geographical and relates to the basic idea that locations close together in space often have similar values of outcome variables while locations far apart are often different. This spatial autocorrelation must be allowed for in spatial analyses. This may have an impact on the structure and form of the likelihood models that are assumed for spatial data. The assumption made in the construction of conventional likelihoods is that the individual contribution to the likelihood is independent and this independence allows the likelihood to be derived as a product of probabilities.

### 3.2.5 INTEGRATED NESTED LAPLACE APPROXIMATION (INLA)

INLA provides a fast and yet quite exact approach to fitting latent Gaussian models which comprise many statistical models, including models with temporal or spatial dependence structures. As a result, many complex models that previously required the use of time-consuming MCMC calculations can be fitted fast and conveniently. Log Gaussian Cox processes, a particularly flexible class of spatial point process models are a special case of latent Gaussian models. A complex point process models, including hierarchically marked point processes may conveniently be fitted with INLA. Standard approaches to parameter estimation for complex models based on MCMC, for example, would be very cumbersome and computationally prohibitive. Fitting spatial point process models to some spatial patterns is computationally intensive due to – amongst other things – the large number of individual points in the data set.

## 3.3 MODEL FITTING IN INLA

A log Gaussian Cox process is a hierarchical Poisson process with random intensity $\Delta_t(s) = \exp(\eta_i(s))$, where $\{\eta_t(s) : s \in \mathrm{R}^d\}$ denotes a Gaussian field. This type of model is a special cases of the more general class of latent Gaussian models, for which deterministic Bayesian inference can be performed using the INLA-methodology.

In general, latent Gaussian models can be described as a subclass of structured additive regression models, in which the predictor can be expressed in terms of linear and non-linear effects of covariates. Explicitly, the mean $\mu_j = E(y_j)$ of observations is linked to a predictor.

$$\eta_j = g(\mu_j) = \beta_0 + \sum_{\alpha} \beta_\alpha Z_{\alpha,j} + \sum_{y} f_y(C_{y,j}) \tag{3.33}$$

where $\beta_0$ denotes an intercept, while the sets $\{\beta_\alpha\}$ and $\{f_\gamma(.)\}$ denote linear effects of covariates $\{z_\alpha\}$ and non-linear effects of covariates $\{c_\gamma\}$, respectively. By assigning Gaussian priors to all random terms in (3.33), we obtain a latent Gaussian model. Here, we fit a log Gaussian Cox process to a specific two-dimensional point pattern $x_t$ discretising the observation window $S$ into $N$ grid cells $\{s_i\}_{t=1}^{N}$ where each cell has area $|s_i|$. For each time point $t = 1,...,T$, let $y_{ti}$ denote the observed number of points in grid cell $s_i$. Conditional on the intensities $\Delta_t(s) = \exp(\eta_i(s))$, the joint pattern of replicates can be described as a superposition of realizations from independent Poisson processes, that is,

$$y_{ti} \mid n_t(s_i) \sim Poisson(|s_i| \exp(\eta_t(s_t))). \tag{3.34}$$

We assume that the log-intensity of the Poisson processes can be described by the linear predictor

$$\eta_t(s_i) = \beta_0 + \beta_t + \sum_{\alpha} \beta_\alpha z_{t\alpha}(s_i) + \sum_{\gamma} f_\gamma(c_{t\gamma}(s_i)), \tag{3.35}$$

in which the off -set $\beta_0$ represents an intercept common to all time points while the factor $\beta_t$ accounts for variation in intensity across different time points. As in (3.33), the set $\{\beta_\alpha\}$ accounts for linear effects of covariates $\{z_{t\alpha}(.)\}$, which may or may not vary with time. We also include potentially smooth effects of covariates $\{c_{t\gamma}(.)\}$, in which the functions $\{f_\gamma(.)\}$ are estimated based on all replicates.

The primary aim in using the INLA-methodology is to find posterior estimates of all the random terms in the log-intensity in (3.35), numerically. These terms are collected in a latent field, $\zeta = \{\beta_0, \beta_t, \{\beta_\alpha\}, \{f_y(.)\}\}$, and assigned Gaussian priors such that the resulting model can be viewed as a latent Gaussian model.

We shall consider the problem of modeling the distribution of a set of continuous variables $x_1,...,x_d$ which we will collectively denote by the vector $\mathbf{x}$. A standard approach to the problem of density distribution estimation involves a parametric model in which a specific form for the density is proposed which contains a number of adaptive parameters. Values for these parameters are then determined from an observed data set $D = \{\mathbf{x}_1,...,\mathbf{x}_N\}$ consisting of N data vectors. The most widely used parametric model is the normal, or Gaussian distribution given by

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) = \left((2\pi)^d |\Sigma|\right)^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\} \qquad (3.36)$$

Where $\boldsymbol{\mu}$ is the mean, $\Sigma$ is the covariance matrix, and $|\Sigma|$ denotes the determinant of $\Sigma$. One technique for setting the values for these parameters is that of maximum likelihood which involves consideration of the log probability of the observed data set given parameters, that is

$$L(\boldsymbol{\mu}, \sum) = \ln p(D \mid \boldsymbol{\mu}, \sum) = \sum_{n=1}^{N} \ln p(\mathbf{x}_n \mid \boldsymbol{\mu}, \sum) \tag{3.37}$$

in which it is assumed that the data vectors $\mathbf{x}_n$ are drawn independently from the distribution. When viewed as a function of $\boldsymbol{\mu}$ and $\sum$, the quantity $p(D|\boldsymbol{\mu}, \sum)$ is called the likelihood function. Maximization of the likelihood with respect to $\boldsymbol{\mu}$ and $\sum$ leads to the set of parameter values which are most likely to have given rise to the observed data set. For the Gaussian distribution (3.36) the log likelihood (3.37) can be maximized analytically, leading to the intuitive result (3.36) that the maximum likelihood solution $\hat{\boldsymbol{\mu}}$ and $\hat{\sum}$ are given by

$$\hat{\boldsymbol{\mu}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathbf{x}_{\mathrm{n}} \tag{3.38}$$

$$\hat{\sum} = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{x}_n - \hat{\boldsymbol{\mu}})(\mathbf{x}_n - \hat{\boldsymbol{\mu}})' \tag{3.39}$$

corresponding to the sample mean and sample covariance respectively.

As an alternative to maximum likelihood, we can define priors over of $\boldsymbol{\mu}$ and $\sum$ using Bayes' theorem, together with the observed data, to determine the posterior distribution. The posterior distribution is given below

$$\begin{aligned} \pi(\boldsymbol{\zeta}, \boldsymbol{\theta} \mid \mathbf{y}) &\propto \pi(\boldsymbol{\zeta}, \boldsymbol{\theta}) \pi(\mathbf{y} \mid \boldsymbol{\zeta}, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\zeta} \mid \boldsymbol{\theta}) \prod_{\mathrm{i}} \pi(\mathrm{y}_{\mathrm{i}} \mid \zeta_{\mathrm{i}}, \boldsymbol{\theta}) \end{aligned} \tag{3.40}$$

.

The posterior marginals of each element $j$ of the latent field can be expressed by

$$\pi(\zeta_j \mid \mathbf{y}) = \int \pi(\zeta_j \mid \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} \tag{3.41}$$

where the vector denotes the hyper-parameters of the model. Here, (3.41) includes the parameters used in defining prior distributions for the precision (inverse variance) of the Gaussian priors, in which

$$\pi(\theta_j \mid \mathbf{y}) = \int \pi(\boldsymbol{\theta} \mid \mathbf{y}) d\theta_{-j} \tag{3.42}$$

where equation (3.42) is the posterior marginals of the hyper- parameters.

Applying the INLA-methodology, the marginals in (3.41), (3.42) are estimated combining analytical approximations with numerical integration. The first step in this procedure is to estimate $\pi(\boldsymbol{\theta} \mid \mathbf{y})$ using Laplace approximation. The Laplace approximation is given by

$$\tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \left. \frac{\pi(\boldsymbol{\zeta}, \boldsymbol{\theta} \mid \mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \mathbf{y}} \right| \boldsymbol{\zeta} = \boldsymbol{\zeta}^*(\boldsymbol{\theta}) \tag{3.43}$$

where the full conditional $\pi(\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \mathbf{y})$ of the latent field is approximated by a Gaussian distribution $\tilde{\pi}_G(\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \mathbf{y})$, evaluated at the mode $\boldsymbol{\zeta}^*(\boldsymbol{\theta})$. Secondly, estimates of the marginals $\pi(\zeta_j \mid \boldsymbol{\theta}, \mathbf{y})$ in (3.41) can be found either using a Laplace or a simplified Laplace approximation. Alternatively, the marginals can be estimated using a Gaussian approximation derived from $\tilde{\pi}_G(\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \mathbf{y})$. Although the Gaussian approximation might provide some inaccuracies in estimating the marginals, this approach is used here to speed up calculations. The third and final step in approximating the marginals of the latent field is to use numerical integration with respect to $\boldsymbol{\theta}$. In order to do this the Laplace approximation $\tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y})$ is explored numerically to find support points for the numerical integration. The resulting approximation to (3.41) is given by

$$\tilde{\pi}(\zeta_j \mid y) = \sum_k \tilde{\pi}_G(\zeta_j \mid \theta_k, y)\tilde{\pi}(\theta_k \mid y)\Delta_k, \qquad (3.44)$$

Where $\tilde{\pi}(\theta_k \mid \mathbf{y})$ is found by numerical integration in (3.42), $\tilde{\pi}_G(\zeta_{jj} \mid \theta_k, \mathbf{y})$ is the Gaussian approximation derived from $\tilde{\pi}_G(\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \mathbf{y})$ and $\Delta_k$ denotes the area weight corresponding to integration point $\theta_k$. The resulting approximation has been shown to be very accurate (Rue et al. 2009) and is also computationally more efficient than using MCMC- approaches.

**Poisson regression**

Let $Y_{ilk}$ be the total number of TB cases observed in $k^{th}$ age group, $j^{th}$ urban residents, $i^{th}$ residential district , and $N_{ijk}$ the corresponding population at risk. Then $Y_{ijk} \sim Poisson(\mu_{ijk})$, given all the random effects. The Poisson models for hospital TB tested patients has the form

$$\log(\mu_{ijk}) = \log(N_{ijk}) + \beta_0 + \beta_1 x_i + \beta_2 x_j + s_i + \varepsilon_{ijk} \qquad (3.45)$$

where $\log(N_{ijk})$ is an offset, $s_i$ and $\varepsilon_{ijk}$ are random effects that allow for spatially structured variation and unstructured heterogeneity respectively.

**Inference model**

For the spatial model, a Bayesian approach was used for inference. The following prior distribution was specified for the parameters in (3.45). A conditional autoregressive (CAR) prior was chosen to model the spatial autocorrelation effects; the unstructured heterogeneity component was assigned a zero mean Gaussian process with heterogeneity variance $\tau_\varepsilon^{\,2}$. The fixed effects $\boldsymbol{\beta}$ were assigned diffuse priors. The calendar effect was assigned second order random walk priors for flexible smoothing, that is,

$$f_t \mid f_{-t} \sim N(2f_{t+1} - f_{t+2}, {\tau_f}^2) \text{ for } t = 1, \tag{3.46}$$

where $\tau_f^2$ is a smoothing variance and $f_{-t}$ denote all elements of **f** except $f_t$ The variance components $(\tau_s^2, \tau_e^2, \tau_f^2)$ are assumed to follow an inverse Gamma with parameters 0.001 and 0.001. The Bayesian model was implemented in R-INLA.

**Issues of spatial Scale –Prior Choice**

In an analysis of a spatial pattern, it is crucial to bear in mind the spatial scales that are relevant for a specific spatial data set. Here, we assume that social behaviour among the individual operates at a local spatial scale and that the association with environmental covariates operates on the scale of the variation in these covariates and hence often on a larger spatial scale. The large-scale spatial effect in (3.45) is included as a spatially structured error term to account for any spatial autocorrelation unexplained by covariates in the model. The choice of the inverse gamma prior for the precision of the spatially structured effect determines the smoothness of the spatial effect and, through this, the spatial scale at which it operates. To avoid overfitting, and in order to obtain a model describing a generally interpretable trend we choose the prior so that the spatial effect operates at a similar spatial scale as the covariate. This ensures that the spatially structured effect does not operate on a smaller scale than the covariate as it would otherwise be likely to explain the data better than the covariates, rendering the model rather pointless. We approach this by repeatedly fitting a simple model.

If small scale inter-individual spatial behaviour is of specific interest in an application it may be modelled by the constructed covariate to account for local spatial behaviour. However, the approach we take here is currently the only way to incorporate local structures into a complex model that may be fitted with INLA. The choice of grid size is also linked to issues of spatial scale. Here we apply the same grid and grid resolution as used in the data collection of locations of the TB Patients in Kenya.

### 3.4 ANALYSIS SOFTWARE

Analysis software used in the study is R-INLA. The Integrated Nested Laplace Approximation (INLA) approach proposed by Rue et al., 2009, is a computationally effective alternative to MCMC for Bayesian inference. INLA is designed for latent Gaussian models; a very wide and flexible class of models ranging from (generalized) linear mixed to spatial and spatio-temporal models. Combined with the Stochastic Partial Differential Equation (SPDE) approach (Lindgren et al., 2011); one can accommodate all kinds of geographically referenced data, including areal and geostatistical ones, as well as spatial point process data. The implementation interface covers stationary spatial models, non-stationary spatial models, and also spatio-temporal models, and is applicable in epidemiology, ecology, environmental risk assessment, as well as general geostatistics

# CHAPTER 4: DATA ANALYSIS AND RESULTS

## 4.1 DESCRIPTIVE ANALYSIS

**Table 4.1: TB prevalence per district**

| Prevalence level( Old Districts) | | | |
|---|---|---|---|
| District | Proportion | District | Proportion |
| Nandi | 0.010391832 | Moyale | 0.000000486 |
| Turkana | 0.010703612 | Isiolo | 0.00000178 |
| Kirinyaga | 0.010711705 | Marsabit | 0.000001784 |
| Tana River | 0.010864852 | Tharaka | 0.00000194 |
| Buret | 0.011353426 | Keiyo | 0.00000243 |
| Gucha | 0.011525033 | Koibatek | 0.0000034 |
| Samburu | 0.011672984 | Trans Mara | 0.00000551 |
| Kakamega | 0.014444989 | Baringo | 0.00000567 |
| Suba | 0.014750608 | Embu | 0.000006004 |
| Taita Taveta | 0.015570415 | West Pokot | 0.00000632 |
| Makueni | 0.015579164 | Mt. Elgon | 0.00000665 |
| Lugari | 0.015888511 | Mandera | 0.000011178 |
| Machakos | 0.015903258 | Garissa | 0.000014426 |
| Butere Mumias | 0.016386006 | Wajir | 0.00002268 |
| Mwingi | 0.020102577 | Mbeere | 0.000650301 |
| Bondo | 0.020586128 | Laikipia | 0.002275805 |
| Trans Nzoia | 0.021078919 | Narok | 0.002600065 |
| Nyando | 0.023342246 | Muranga | 0.002601519 |
| Bungoma | 0.023856268 | Maragua | 0.002929506 |
| Uasin Gishu | 0.023998547 | Kilifi | 0.00293339 |

| | | | |
|---|---|---|---|
| Vihiga | 0.024491003 | Teso | 0.003246796 |
| Nyandarua | 0.025628422 | Meru South | 0.003732298 |
| Nakuru | 0.026117004 | Meru Central | 0.003903568 |
| Busia | 0.027563761 | Kuria | 0.004383249 |
| Kisii North | 0.033239023 | Lamu | 0.004863879 |
| Kisumu | 0.033880084 | Kajiado | 0.005031601 |
| Homa Bay | 0.035822233 | Kitui | 0.006818504 |
| Rachuonyo | 0.037442713 | Bomet | 0.00714049 |
| Migori | 0.043927888 | Malindi | 0.007789497 |
| Nyeri | 0.046379191 | Kericho | 0.008115546 |
| Mombasa | 0.054002914 | Marakwet | 0.008269808 |
| District | Proportion | District | Proportion |
| Siaya | 0.062730515 | Central Kisii | 0.009415809 |
| Kiambu | 0.065024314 | Thika | 0.0097438 |
| Nairobi | 0.170261384 | Meru North | 0.009904871 |
| Kwale | 0.010225896 | | |

**Table 4.2: TB prevalence per province**

| Province | Proportion |
|---|---|
| Central | 0.163018457 |
| Coast | 0.106250843 |
| Eastern | 0.076606535 |
| Nairobi | 0.170261384 |
| North Eastern | 0.000048284 |
| Nyanza | 0.331045529 |
| Rift Valley | 0.148772969 |
| Western | 0.125883984 |

Table 4.1 and Table 4. 2 display a summary of the proportion of residents with TB. Nyanza Province has a prevalence of 0.33 of the total population while North Eastern Province has 0.00048284 of the total population infected with TB.

Table 4.3 shows a summary of Health Facilities in Kenya. The composition and distribution of health facilities across the country shows that Nairobi as a district has the highest number of health facilities followed by Nakuru and Machakos respectively. The districts with the least number of health facilities include; Moyale and Mtelgon with a total count of 21 health facilities.

Figure 4.1 presents a summary of the Health Facility per Province .The province with the largest number of health facilities is Rift Valley followed by Eastern and Nyanza Province respectively.

**Table 4.3: Health facilities in Kenya**

| Health Facilities In Kenya | | | |
|---|---|---|---|
| District(Old) | Number of Facilities | District(Old) | Number of Facilities |
| BARINGO | 120 | NAIROBI | 272 |
| BOMET | 69 | NAKURU | 202 |
| BONDO | 50 | NANDI | 126 |
| BUNGOMA | 77 | NAROK | 66 |
| BURET | 70 | NYAMIRA | 106 |
| BUSIA | 48 | NYANDARUA | 69 |
| BUTERE/MUMIAS | 60 | NYANDO | 48 |
| CENTRAL KISII | 67 | NYERI | 146 |
| EMBU | 67 | RACHUONYO | 61 |
| GARISSA | 66 | SAMBURU | 49 |
| GUCHA | 56 | SIAYA | 83 |
| HOMA BAY | 53 | SUBA | 43 |
| ISIOLO | 46 | TAITA TAVETA | 64 |
| KAJIADO | 119 | TANA RIVER | 79 |
| KAKAMEGA | 80 | TESO | 26 |
| KEIYO | 59 | THARAKA | 22 |
| KERICHO | 86 | THIKA | 113 |

| Health Facilities In Kenya | | | |
|---|---|---|---|
| District(Old) | Number of Facilities | District(Old) | Number of Facilities |
| KIAMBU | 98 | TRANS MARA | 82 |
| KILIFI | 53 | TRANS NZOIA | 59 |
| KIRINYAGA | 100 | TURKANA | 103 |
| KISUMU | 79 | UASIN GISHU | 145 |
| KITUI | 133 | VIHIGA | 54 |
| KOIBATEK | 44 | WAJIR | 58 |
| KURIA | 44 | WEST POKOT | 61 |
| KWALE | 67 | MARSABIT | 40 |
| LAIKIPIA | 87 | MBEERE | 43 |
| LAMU | 30 | MERU CENTRAL | 101 |
| LUGARI | 37 | MERU NORTH | 69 |
| MACHAKOS | 147 | MERU SOUTH | 62 |
| MAKUENI | 143 | MIGORI | 109 |
| MALINDI | 46 | MOMBASA | 93 |
| MANDERA | 40 | MOYALE | 21 |
| MARAGUA | 57 | MT ELGON | 21 |
| MARAKWET | 67 | MURANGA | 68 |
| MWINGI | 99 | | |

**(Source Kenya National Health Facility Database 2008)**

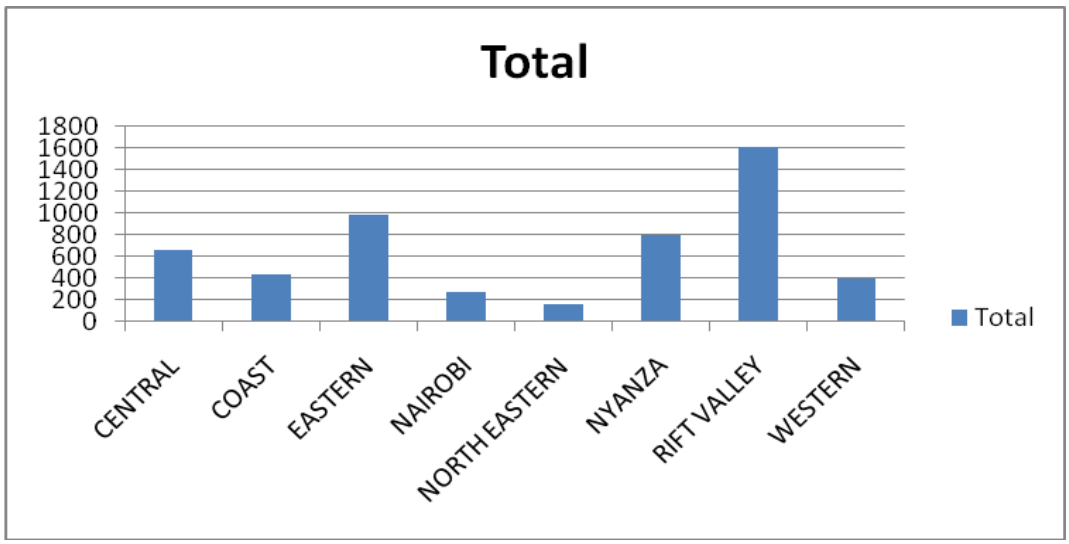


**Figure 4.1: Summary of health facilities in Kenya**

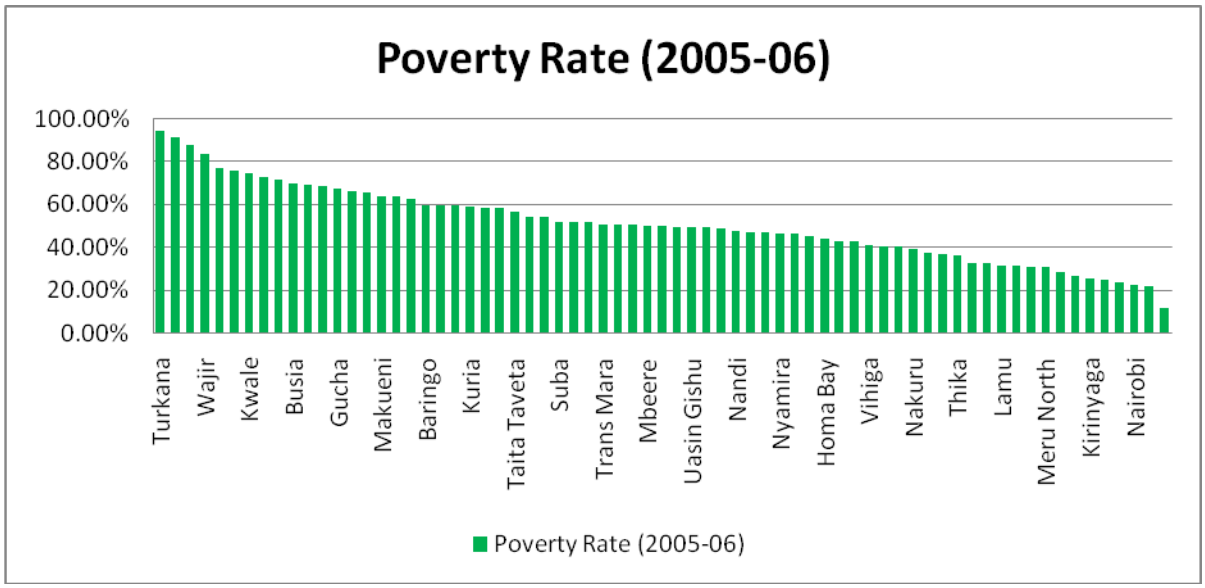


**Figure 4.2: Poverty rate in Kenya per district**

Figure 4.2 gives a summary of the poor district with the leading district being Turkana followed by Wajir and Kwale respectively. Proportion of poor in the respective district will be a covariate in our model to test if it is significant or otherwise.

## 4.2 INLA ANALYSIS

**Results from INLA**

1. We first fitted a Bayesian logistic regression and below are the summary results.

**Table 4.4: Results from INLA-R**

| Fixed Effects | Mean | SD | 0.025 quant | 0.5 quant | 0.975 quant | kld |
|---|---|---|---|---|---|---|
| Intercept | -0.958 | 0.553 | -2.065 | -0.951 | 0.106 | 2.596e-04 |
| Urban | 0.007 | 0.004 | -0.001 | 0.007 | 0.015 | 2.294e-05 |
| Child taught | 0.040 | 0.012 | 0.017 | 0.040 | 0.063 | 3.217e-05 |
| Propoor | -0.164 | 0.254 | -0.664 | -0.164 | 0.333 | 1.745e-05 |
| TB positive taught | -0.015 | 0.024 | -0.062 | -0.015 | 0.031 | 4.314e-04 |
| Prop taught | 0.832 | 0.770 | -0.655 | 0.823 | 2.370 | 4.314e-04 |
| Tested for HIV | -0.015 | 0.019 | -0.052 | -0.015 | 0.021 | 1.471e-04 |
| Age 25-49 Years | -0.008 | 0.018 | -0.043 | -0.008 | 0.028 | 1.8072e-04 |
| Age 15-24 Years | 0.708 | 0.052 | 0.605 | 0.708 | 0.808 | 1.495e-06 |

The model has no random effects

The model has no hyperparameters

Expected number of effective parameters (stddev): 9.006(0.00)

Number of equivalent replicates 44.30

Deviance Information Criterion: 996.48

Effective number of parameters: 8.973

Marginal likelihood:-547.86

Posterior marginals for linear predictor and fitted values computed

2. We then fitted a Bayesian Logistic regression with additional random effects (geo locations) below are the results.

**Table 4. 5: Results from INLA -R with random effects**

| Fixed Effects | Mean | SD | 0.025 quant | 0.5 quant | 0.975 quant | kld |
|---|---|---|---|---|---|---|
| Intercept | -0.678 | 0.572 | -1.819 | -0.671 | 0.425 | 5.341e-04 |
| Urban | 0.007 | 0.004 | -0.002 | 0.007 | 0.015 | 2.864e-05 |
| Child taught | 0.028 | 0.013 | 0.003 | 0.028 | 0.053 | 1.012e-04 |
| Propoor | -0.400 | 0.279 | -0.950 | -0.399 | 0.144 | 3.766e-06 |
| TB positive taught | -0.006 | 0.024 | -0.054 | -0.006 | 0.042 | 3.822e-04 |
| Prop taught | 0.408 | 0.802 | -1.145 | 0.401 | 2.004 | 5.227e-04 |
| Tested for HIV | -0.010 | 0.019 | -0.047 | -0.0101 | 0.026 | 6.741e-05 |
| Age 25-49 Years | -0.006 | 0.018 | -0.041 | -0.006 | 0.030 | 1.209e-04 |
| Age 15-24 Years | 0.641 | 0.060 | 0.522 | 0.641 | 0.758 | 6.427e-06 |

Random effects:

Maximum KLD

Longitude and latitude give a random walk 1 model.

Model hyperparameters:

Table 3.1

**Table 4.6: Random effects**

| | Mean | SD | 0.025 quant | 0.5 quant | 0.975 quant |
|---|---|---|---|---|---|
| Precision for latitude | 18656.736 | 18519.282 | 1266.764 | 13185.610 | 67556.664 |
| Precision for longitude | 141.184 | 340.562 | 6.676 | 56.852 | 799.836 |

Expected number of effective parameters (stddev): 12.08(1.85)

Number of equivalent replicates: 33.03

Deviance Information Criterion (DCI): 985.25

Effective number of parameters: 13.07

Marginal Likelihood: -1535.71

Posterior marginals for linear predictor and fitted values computed

From the two results the model, the second one with spatial effect is more efficient with a DIC of 985.25 as compared to the model with no spatial effects DIC of 996.48. And hence the model with additional spatial effect is more precise.

**Interpretation of the results**

For the random effects precision for latitude with mean of 18656.736 and SD 0f 18519.282 lies between 0.5 quant and 0.975 quant which implies that the latitude effect is highly significant to the distribution of TB epidemic in Kenya.

The other random effect longitude with mean of 141.184 and a SD of 340.562. The mean lies between 0.5 quant and 0.975 which implies that it is highly significant it explains over 50% and just below 97%. The model yields a maximum KLD-Kullback Leibler Deviance between the Gaussian and the simplified laplace approximation to the marginal posterior density.

Fixed effects; The intercept has a mean of -0.677839790 and a SD of 0.571899092 .The mean lies slightly above 0.5 quant and yields a kld of 5.340830e-04.The intercept is significant at just about 50%.

The effect on urban resident has a mean of 0.006515052 and a SD of 0.004254495 and yields a KLD of 2.864264e-05 . The mean for the Urban resident effects to patients with TB in Kenya lies in the 0.5 quant which implies that the effect is significant at 50%. The effect on if children have been taught about TB and how it could be controlled has a mean of 0.027931627 and a SD of 0.012829047 which has a kld of 1.011701e-04 .The effect is significant and lies in the 0.5 quant.

The effect on the proportion poor per district has a mean of -0.400422160 and a SD of 0.279144267 which has kld of 3.765746e-06. The mean lies approximately in the 0.5 quant which implies that it is significant.

The effect on the patients with TB being taught has a mean of -0.005741851 and a SD of 0.024453863 which has KLD of 3.822112e-04.The mean lies in the 0.5 quant which implies that it is significant at 50%

The effect the total proportion of the population taught has a mean of 0.408466779 and a SD of 0.802411793 with a KLD of 5.226588e-04. The mean lies just above the 0.5 quant which implies that it is significant.

The effect on patients tested for HIV has a mean of -0.010080341 and a SD of 0.018805976 with a KLD of 6.741331e-05 . The mean does not lie in any of the 3 quantrants and hence it is not significant.

The age effect on age (25-49 years) has a mean of -0.005848474 and a SD of 0.018238842 with a KLD of 1.209031e-04.the mean lies in the 0.5 quant which implies that it is significant.

The other age group of (15-24 years) has a mean of 0.640508475 and a SD of 0.060348362 with KLD of 6.426770e-06. The mean lies in the 0.5 quant and this implies that it is significant.

The age between 15 to 24 years has a higher mean than the age between 25 to 49 years , this implies that persons between age 15 to 24 have a high likelihood of being infected with TB as compared to persons aged 25 to 49 years.

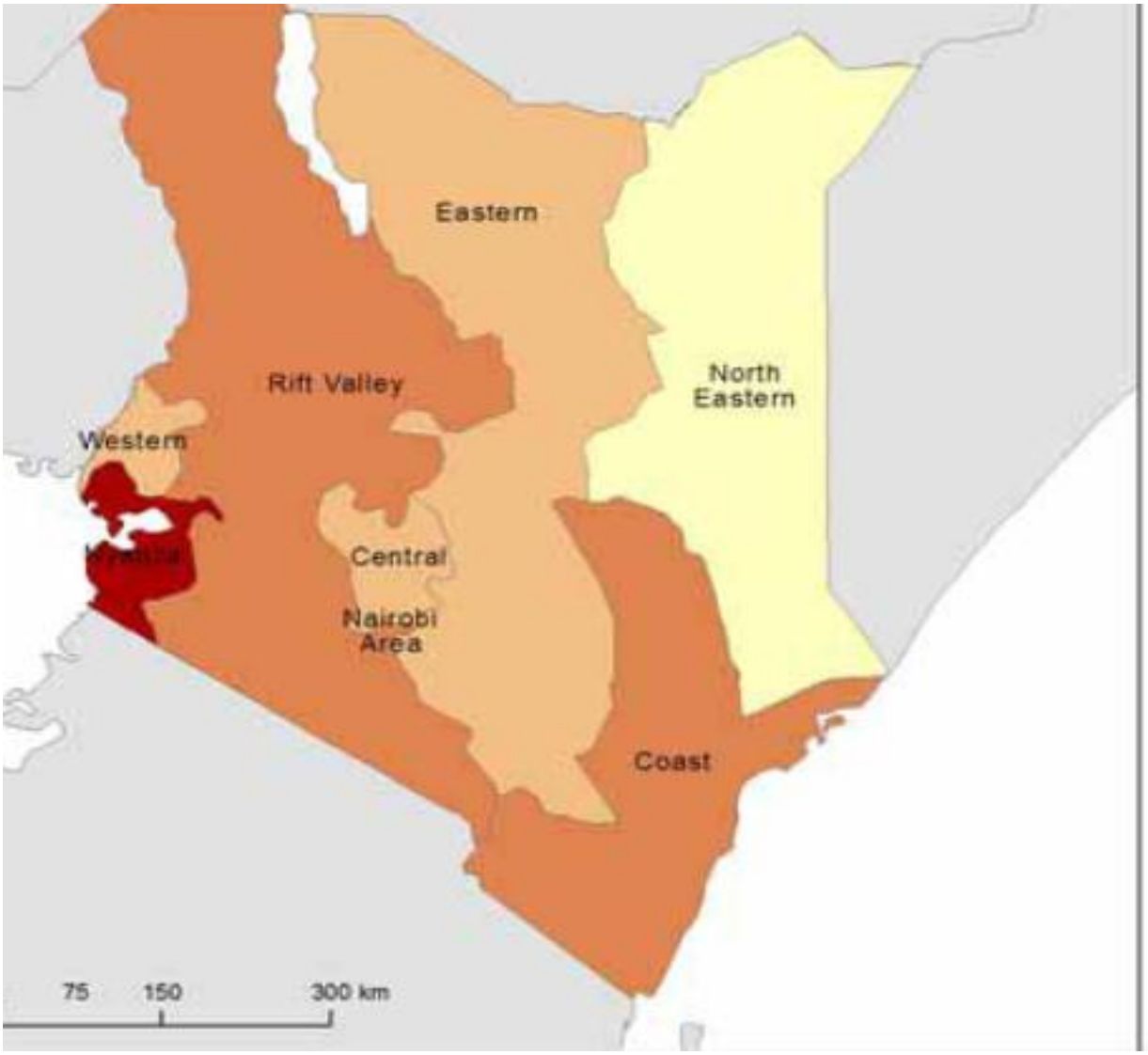


**Figure 4. 3: Distribution of TB prevalence**

Key:

–less than 2 %( North Eastern)

–2- 5 %( Eastern, Central ,Nairobi, Western)

–5-10% (Coast)

–l0-15% (Rift Valley)

More than 15% (Nyanza)

Figure 4.3, illustrates that Nyanza province has the highest population in Kenya with TB prevalence at over 15%, this is followed by Rift valley with a prevalence of between 10% to 15 %, followed by coast province with a prevalence of 5% to 10 %. The province with the least

number of TB infections is North Eastern with less than 2% followed by Eastern, Central Nairobi and Western provinces all with a prevalence of between 2% and 5 %.

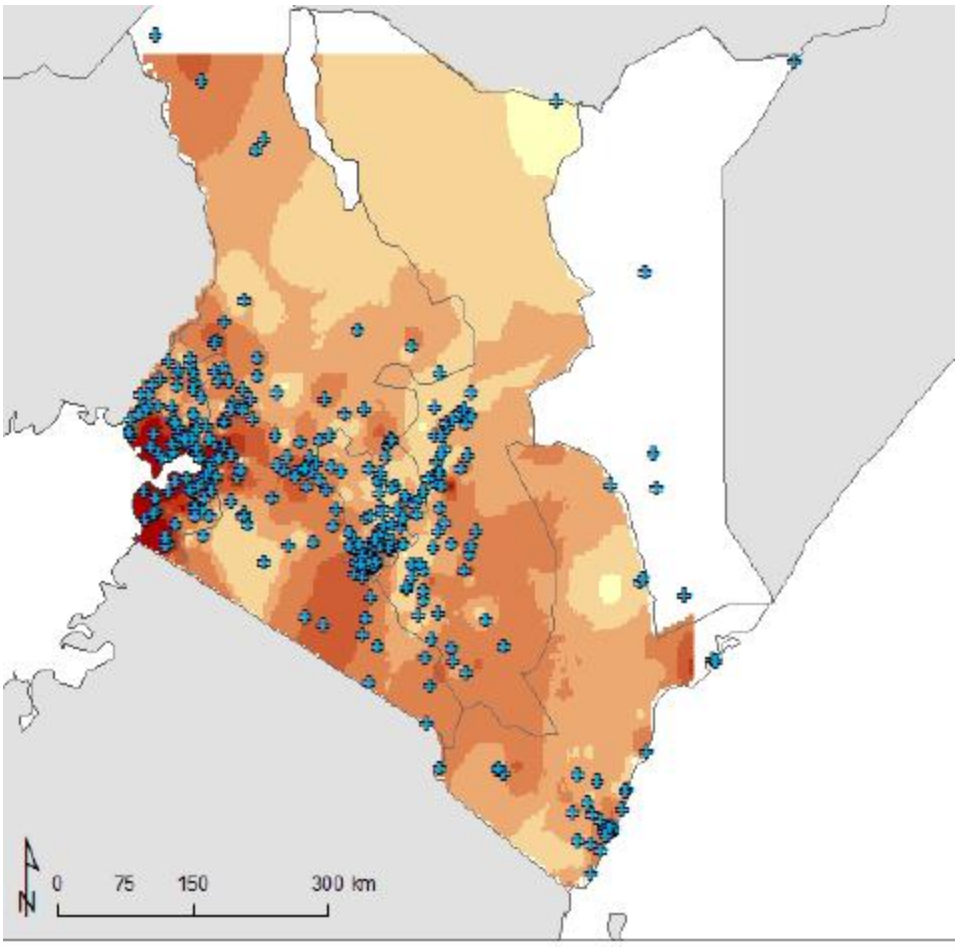


**Figure 4.4: Distribution of health facilities in Kenya**

Figure 4.4 illustratess the distribution of health facilities in Kenya as per the Kenya National health facility database 2008. From figure 4.4 entire of North Easter province has fewer facilities has compared to other regions in the country.

The distribution is skewed towards the major cities and highly productive parts of the country with North Eastern province have a negligible number as well as the bounder to South Sudan

# CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

## 5.1 CONCLUSIONS

Urban residence effect is significant and lies in the 0.5 quant. This could be the fact that most of the urban centres in Kenya for example Nairobi has a larger population living in slum areas where ventilation is poor and hence could cause the infection and spread of TB.

The proportion poor effect is significant, the mean lies in the 0.5 quant and this implies that poverty could greatly influence the infection and spread of TB in Kenya. The effect of Age (25-49 years) is significant and lies in the 0.5 quant. This implies that TB is most likely to affect the active members of the society this is as well confirmed by the result found for age (15-24 Years) Most of the population infected by TB lie in 15 to 24 years age group which has a higher mean. .The effects on; children taught, TB infected persons taught on the infection and spread of TB and control are significant and falls in the 0.5 quant. This implies that if more people are knowledgeable on the control and spread of communicable diseases the less the burden of the disease

The total population taught is significant. This implies that if the total population in various districts is well informed on measure to be taken in the control of TB the less the number of TB cases in Kenya.

The effect on the patients tested for HIV is not significant and this implies that testing for HIV does not have any effects on TB prevalence in Kenya as per the data collected and analyzed

Figure 4.3 shows that Nyanza province has the largest proportion of people with TB of more than 15%; this province is followed by Rift Valley province with prevalence of 10% to 15%, followed by Coast province with a prevalence of 5% to 10%. The province with the least number of TB incidences is North Eastern province with a prevalence of less than 2%. This implies that TB is clustered and different geographical regions have different prevalence level

Mapping the distribution of health facilities in the country, from figure 4.4 we find that the spread of health facilities is not evenly distributed with North Eastern province having a

negligible number of health facilities. This implies that most of the TB infected cases in the regions where there are few or no health facilities could fail to be represented and hence the regions with few health facilities may tend to report few events of TB infections as to compared to the regions with more health facilities. Hence the government should ensure proportional distribution of health facilities to be able to register all the patients with TB

From figure 4.2 shows a graphical representation of poverty distribution per district 2005-2006. The poorest district is Turkana and the richest Nairobi.

The random effects precision for latitude and longitude is significant and lies between 0.5 quant and 0.975 quant., this implies that TB prevalence varies significantly within different geographical location.

**5.2 RECOMMEDATIONS**

The government should ensure that there is a continuous education on the infection, spread and control of TB at all levels (primary schools, secondary schools, universities, churches and other social gatherings). This will ensure that the population is well informed on the different ways they could get tested on time and be cured.

Poverty is a great contributor of disease burden; and hence the government should put measures in place to upgrade the standards of living and encourage the citizens to participate in initiative which will improve their per capita income.

For there to be good and proper reporting of TB cases, the government should ensure that health facilities are proportional to the population per a certain region.

**REFERENCES**

Burslem, D.F.R.P., Garwood, N.C. and Thomas, S.C. (2001). Tropical forest diversity – the plot thickens. *Science,* **291**: 606–607.

Clayton, D. and Kaldor, J. (1987). Emphirical Bayesian estimates of age-standardized relative risks in desease mapping. *Biometrics,* **43**:671-681

Cressie, N. and Chan, N.H. (1989). Spatial modeling of regional variables. *Journal of the American Association,* **84**:393-401

Cressie, N.A.C. (1991). *Statistics for spatial data*. John Wiley and Sons, New York.

Draper, D. 2000. *Bayesian Hierarchical Modelling.* Draft version can be found on the web at HTTP://www.bath.ac.uk/~masdd/

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *J Am. Stat. Asso.,* **85:** 398-409.

Geman, S. and Geman. D. (1984). Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEE Transactions on Pattern Analysis and Machine Intelligence, **6**:721-*741.

Gilks W.R., Richardson, S. and Speigelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice.* London : Chapman and Hall.

Illian, J.B. and Rue, H. (2010) A toolbox for fitting complex spatial point process models using integrated Laplace transformation (INLA). *Technical Report, Trondheim University*.

Illian, J.B. and Hendrichsen, D.K. (2010). Gibbs point process models with mixed effects. *Environmentrics,* **21**: 341–353.

Illian, J.B., Sørbye, S.H. and Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *To appear in Annals of Applied Statistics.*

Isaaks, E.H. and Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics.* Oxford University Press, New York.

Snow, M.D. (1854).*The Mode of Communication of Cholera*. John Churchill, New Burlington Street, England.

Journel, A.G. and Huijbregs, C.J. (1978) *Mining Geostatistics*. Academic Press.

Kazembe, L.N., Kleinschmidt, I., Holtz, T.H. and Sharp, B.L. (2006). Spatial analysis and mapping of malaria risk in Malawi using point referenced prevalence of infection data. *International Journal of health geographics,* **5**:41

Kenya National Bureau of Statistics (KNBS) (2010).2009 *Kenya Population and Housing Census.* Ministry of Planning and Development Government of Kenya, Kenya.

Law, R., Illian, J.B., Burslem, D.F.R.P., Gratzer, G., Gunatilleke, C.V.S. and Gunatilleke, I.A.U.N. (2009). Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology,* **97**:616–628.

Lindgren, F., Rue, H.and Lindstr¨om, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion)." *to appear in Journal of the Royal Statistical Society.*

MacNab, Y. C. (2003). Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease. *Statist. Med*., **22**: 1761–1773.

Marshall, R.J. (1990). A review of the methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society,* **154**:3.

Marten, B. (1960). *Spatial variation.* Springer, New York

Matheron, G. (1965). *Les Variables regionalisees et Leur Estimation.* Paris:Masson.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. *J Amer. Statist. Assoc.*, **44:**335-341.

Metropolis, N., Rosenbluth, A.W., Teller, A. and Teller, H. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics,* **21**:1087-1091.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society,* **135**: 370-384.

Openshaw, S. ( 1984). *The Modifiable Areal Unit Problem*. Geobooks, Norwich, England.

R Development Core Team: R: A language and environment for statistical computing 2004 [http://www.r-project.org.]. R Foundation for Statistical Computing Vienna Austria.

Ripley, B. D. (1981). *Spatial statistics*. John Wiley & Sons, New York.

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*, **71**: 319–392.

Smith, A.F.M. (1991). Bayesian computational methods. *Phil.Trans.R.Soc.Lond.*A, **337**:369-386.

Tanner, M.A. (1996). *Tools for statistical inference*, 3rd ed. Springer-Verlag, New York.

The Global Fund to Fight AIDS, Tuberculosis and Malaria [http://www.theglobalfund.org].

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statistics*, **22**:1701-1762.

Waagepetersen, R. and Guan, Y. (2011). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society*, **71**: 685–702.

Waagepetersen, R.P. (2007) An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, **63**: 252–258.

WHO 1997.Press Release WHO/TB Cap 200.

Wilkinson, P., Thakrar, B., Shaddick, G., Stevenson, S., Pattenden, S. and Landon M. (1997). Cancer incidence and mortality around the Pan Britannica Industries pesticide factory, Waltham Abbey. *Occup Environ Med*., **54**:101–107