Spatio-temporal Characteristics of Tuberculosis in Ghana

Abdul-Karim Iddrisu (abdul-karim.iddrisu@uenr.edu.gh)
University of Energy and Natural Resources
Emmanuel A. Amikiya
Ghana Institute of Management and Public Administration
Francis Kwame Bukari
University of Energy and Natural Resources

Research Article

Keywords: Bayesian spatial and space-time models, Tuberculosis relative risk, baseline predictors, and TB hot-spots.

Posted Date: October 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-802327/v1

Abdul-Karim Iddrisu*,[1], Emmanuel A. Amikiya[2] and Francis Kwame Bukari[1] [1]Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana. [2]Department of Management Science, Ghana Institute of Management and Public Administration,Ghana.

BMC INFECTIOUS DISEASES OPEN ACCESS

Spatio-temporal characteristics of Tuberculosis in Ghana

Abdul-Karim Iddrisu*,[1], Emmanuel A. Amikiya[2] and Francis Kwame Bukari[1]
[1]Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana.
[2]Department of Management Science, Ghana Institute of Management and Public Administration,Ghana.

Abstract

Background: Recent global reports show that the number of Tuberculosis (TB) cases or deaths is declining, however, the rate of decline is not adequate to meet the World Health Organization's (WHO's) mitigation. TB remains a public health problem in Ghana with a significant economic and health burden on citizens and health infrastructure.
Aims: Consequently, there is a need for further studies about the disease aimed at accelerating the rate of decline in cases.
Methods: In this article, we study the spatio-temporal characteristics of TB in Ghana, using data obtained from Ghana National Tuberculosis Programme (NTP) for the 10 regions of Ghana, collected over a six-year period. Bayesian spatial and space-time regression models are used to map the risk of TB infections across the nation, in time and space. The study also examines some baseline predictors of TB infections to ascertain their effects on the TB risk.
Results: Our study results showed that hot-spots of TB cases are observed in the Upper East, Upper West, Volta, Western, and Central regions and low risk in the Northern, Ashanti, Greater Accra, Brong Ahafo, Eastern and Western regions. We observed clustering of risk between neighboring regions. TB cure rate, TB success rate, knowledge about TB, awareness that TB is airborne, HIV prevalence, percentage of literacy, high income are important predictors of TB detection across the 10 regions of Ghana.
Conclusions: Most regions in Ghana have similar TB risk. Efforts for more TB cases detection should be encouraged to increase TB success and cure rate which will lead to substantial decrease in TB spread. There is the need for provision of adequate health facilities with easy access to these facilities irrespective of your income status to bridge the gap between TB cases among the poor and the rich. TB cases are expected to grow exponentially in countries with low success and cure rate. Finally, for a substantial TB cases reduction, there is the need to adopt measures that will increase TB cases detection, TB success and cure rates, TB awareness, knowledge about how TB spread as well adequate health facilities with easy access.

Keywords: Bayesian spatial and space-time models, Tuberculosis relative risk, baseline predictors, and TB hot-spots.

1 Background

Tuberculosis (TB) is an infectious disease transmitted by *bacillus Mycobacterium tuberculosis* [1]. It is a disease that spreads across age, gender, race, healthy and sick human populations across the globe [1]. In 2019, approximately 10 million people tested positive for TB with an estimated 1.2 million HIV-negative deaths [1]. Male adults (aging 15 years and above) accounted for 56% of the global infections compared to 32% for females in the same age group. Infections among children accounted for 12% of the total case and about 8.2% of the reported cases were HIV-patients [1]. Africa accounted for 25% of the infections, South-East Asia and Western Pacific accounted for 44% and 18% (respectively) of the total infections in 2019. Eastern Mediterranean, America and Europe accounted for 8.2%, 2.9%, and 2.5% respectively, of the total cases [1].

Correspondence: Abdul-Karim Iddrisu (karim@aims.ac.za)
Department of Mathematics and Statistics, University of Energy and Natural Resources, +233, Sunyani, Ghana

However, the World Health Organization (WHO) has called for the development of TB vaccines [1]. Aside vaccines, there is treatment for TB patients, and reports indicate that about 85% of patients can be cured successfully with a 6-month drug schedule [1]. Available reports indicate that treatment drugs have prevented more than 60 millions deaths from 2000 to 2020 [1]. Cumulatively, there was 9% reduction in cases between 2015 and 2019, and about 2.3% reduction between 2018 and 2019 [1]. Europe has achieved 19% reduction in cases and 31% reduction in deaths between 2015 and 2019. Africa has achieved a 16% reduction in cases and a 19% reduction in deaths between 2015 and 2020 [1]. More statistics on TB can be found in [1–6].

Although, global reports on TB indicate a decreasing trend in cases and deaths, the 2015-2020 reduction targets of the WHO, have not been achieved [1]. Apart from Europe, all the other continents have not been able to meet the acceptable reduction levels. Thus, TB is still one of the top 10 causes of death especially in Africa (ranked above HIV/AIDS) [1]. Hence, there is a need for further studies to be conducted on the dynamics of the disease and on mitigation measures for TB in Africa. Ghana as a developing country in Africa, has been affected by the respiratory disease and currently has challenges eradicating TB. The country implemented policies called *Directly Observe Therapy* (DOT) and *National Tuberculosis Programmes* (NTPs) in 1994, to detect and treat TB [1, 7–9]. The implementation of the NTP led to 100% DOTs coverage in 2005 with more TB cases detected for treatment every year after. For instance, TB cases detected increased from 7,425 in 1996 to 15,286 in 2009 [10, 11]. Detailed discussion of the statistics of TB in Ghana can be found in [1, 8–13].

Although TB cases and deaths have declined due to the implementation of mitigation/treatment strategies, TB still remains a life-threatening disease and poses a burden on health infrastructure in Ghana. Hence, TB has gained considerable attention as a topic of research among researchers from diverse backgrounds. The authors in [8, 9, 14] have studied the dynamics of TB indicators as well as risk factors of TB in Ghana. Osei et. al. [8, 9] studied trends of TB detection and treatment outcomes using the logistic regression to assess the relationship between patients and disease characteristics. Further, Osei et. al. [8, 9] have studied TB detection, mortality and co-infection with HIV, using patients data collected in the Volta Region from 2012–2016. The authors used simple and multiple logistic regression to investigate determinants of TB mortality in 10 districts of the Volta Region of Ghana. Aryee et. al. [14] have studied the dynamics of TB using Autoregressive Moving Average (ARIMA) methods and TB data recorded by Korle Bu Teaching Hospital from 2008–2017. Iddrisu et. al. [15] have studied the temporal and geographical pattern of TB prevalence in Ghana between 2015 and 2018.

In the discussion in this paper, we extend the literature on TB cases in Ghana, by using Bayesian hierarchical spatial and space-time models to study the relative risk (RR) of TB and associated risk factors across the 10 Regions of Ghana. Hence, the purpose of this study is to model the spatio-temporal risk pattern of TB in Ghana, using Bayesian hierarchical and space-time models discussed in [5, 16–19]. The relative risk pattern of TB will be estimated and used to identify regions in Ghana that are hot-spots. We also check for clustering of risk between regions and examine the variability of TB risks among the regions in both space and time. The rest of the paper is divided into four main sections. We describe the data in Section 2 and space-time methods in Section 3. The experimental results are presented in Section 4 and the conclusion is drawn in Section 5.

2 Description of data

In this section, we describe the TB detection data used as well as some selected baseline potential predictors of TB cases in Ghana.

2.1 TB cases data

In this study we used TB detection data obtained from Ghana Health Service and National Tuberculosis Programme [10]. The data contain information on TB detection from 2009 to 2017, for the 10 old administrative regions of Ghana. These regions include, Ashanti, Brong Ahafo, Central, Eastern, Greater Accra, Northern, Upper East, Upper West, Volta, and Western.

Figure 1 shows the TB trends in the 10 regions from 2008 to 2017. Generally, there is a decrease in TB cases observed in all regions (except Brong Ahafo Region where TB cases increase) of Ghana from 2008 to 2016. It can also be observed that TB cases in Northern and Upper East increased remarkably between 2016 and 2017, whereas cases in Ashanti Region decreased from 50 per 100, 000 population in 2016 to 45 per 100,000 population in 2017. In the Northern Region, TB cases increased from 24 per 100,000 population in 2016 to 52 per 100,000 population in 2017. Further, cases in Upper East Region increased from 53 per 100,000 population in 2016 to 63 per 100,000 population in 2017. However, the changes in TB cases in the other regions are almost horizontal.

Figure 2 shows the trend of total number of TB cases for each region from 2008 to 2017. It shows that the

highest cases recorded was in the Greater Accra Region from 2008 to 2017. It also shows that the lowest was recorded in the Northern Region. Volta and Western Regions are second and third, respectively, with records slightly lower than Greater Accra Region. In addition, Figure 3 shows the trend of total TB cases in each year/period. The figure shows that the highest number of total TB cases was recorded in 2011 while the lowest was in 2016. It can be observed that TB cases decreased slowly from 2011 to 2017 with increments in 2012 and 2017.

Furthermore, the variability of TB cases from 2008 to 2017 have also been presented using box-and-whisker plots in Figure 4. The overlapping box-and-whisker plots imply that there is no variability in cases among the years. The plots show that TB cases were skewed towards larger numbers from the year 2008 to 2014, and skewed towards smaller numbers from 2015 to 2016. Extremely small numbers were observed in 2013, 2014, 2015, and 2017. Variability across the regions have been presented in Figure 5. None-overlap of box-and-whisker plots imply variability between regions. Thus, there is variability in TB cases among the regions since some of the box-and-whisker plots do not overlap. It can also be observed that TB cases in most of the regions are skewed towards larger numbers except Northern and Upper East Regions (especially, the Northern Region with one extremely large value).

2.2 Baseline predictors

Some baseline regional characteristics on the risk of TB infection have been explored in this study. The set of baseline predictors include doctor to population ratio, nurse to population ratio, HIV prevalence, Tuberculosis cure rate, Tuberculosis success rate, wealth quantiles and the proportions of men/women employed, unemployed, educated, and uneducated. Variables also considered in the study include proportions of people who have heard about the TB disease, have knowledge that TB is airborne, knowledge that TB can be cured, and those who believe that TB status should be kept secret. In the data analyses, all the baseline variables are explored to obtain significant predictors of the TB cases.

3 Methods

This section discusses the distribution of TB cases, spatial and space-time models that will be used to estimate the relative risk across the 10 regions of Ghana.

3.1 Hierarchical space model

For spatial TB data, let $y_i, i = 1, \ldots, n$ denote a Poisson random variable with probability mass function defined as $P(y_i \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ is a vector of relative risk parameters for each region. The variable y_i represents total number of TB cases for region i. It follows that the likelihood function for the Poisson variable y_i is defined as:

$$P(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} P(y_i \mid \boldsymbol{\theta}),$$

with assumption that the sample values of $\mathbf{y} = (y_1, \ldots, y_n)'$ given the parameter estimates $\boldsymbol{\theta}$ are independent [20]. Bayesian modeling framework requires prior distribution of the unknown parameters in the likelihood function for the data. The prior distribution represents the current knowledge of the parameters θ before the data y_i are observed [20]. Under the Bayesian framework, all parameters are stochastic and assigned appropriate distributions called prior distributions [20]. Bayesian modeling framework combines the likelihood function for the data and the prior distributions for the parameters resulting in a distribution known as the posterior distribution [2, 3, 5, 20]. The posterior distribution is defined as $P(\boldsymbol{\theta} \mid \mathbf{y})$, (i.e. probability distribution of the parameters given that the data which is proportional to the product of the likelihood function) while the prior distribution is defined as:

$$P(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid \boldsymbol{\theta})P(\boldsymbol{\theta})}{\int_p L(\mathbf{y} \mid \boldsymbol{\theta})P(\boldsymbol{\theta})d_{\boldsymbol{\theta}}} \quad (1)$$

where the denominator of Equation (1) is called the normalizing constant. It has been shown that the posterior distribution can alternatively be written as:

$$P(\boldsymbol{\theta} \mid \mathbf{y}) \propto P(\mathbf{y} \mid \boldsymbol{\theta})P(\boldsymbol{\theta}),$$

from which parameter estimates are drawn using Integrated Nested Laplace Approach [19, 21, 22] proposed by Håvard et. al.[23]. This is an approximate Bayesian inference approach and has become an established alternative to Markov chain Monte Carlo due to its speed and ease of use via the R-INLA package.

3.1.1 Correlated and uncorrelated heterogeneity structures

Clustering and variability of risk are are studied using correlated and uncorrelated structures, respectively. The use of uncorrelated heterogeneity models with gamma or beta prior distributions for estimating the relative risk of a given disease are useful, however, such models have limitations. Andrew [20, P. 82-84] stated that a gamma distributions is restrict the incorporation of covariates into the modeling process. Another

limitation is that such models do not allow the formulation of a simple and adaptable general form of the gamma distribution with spatially correlated parameters [20, P. 82-84]. Also, Wolpert and Ickstadt [24] have given an example of correlated gamma field models that yield poor results [25]. However, Gaussian models permit incorporation of correlated structure (CH) into the modeling process. Further, variability in the data can be modeled as uncorrelated heterogeneity (UH) using a Gaussian prior distribution with a mean zero and risk variance of the disease across the regions. Both correlated and uncorrelated heterogeneities can be incorporated into the model to account for clustering and heterogeneity of risk. These structures are introduced into the modeling through a log-linear term with additive random effects [3, 26].

Besag et. al.[27] have provided the form of the model with CH and UH structures parameterized as follows:

$$\exp\left(x_i'\beta + \phi_i + u_i\right),$$

where $x_i'\beta$ is the fixed effect component, ϕ_i and u_i are the correlated and uncorrelated heterogeneity components, respectively with separate prior distributions. Often, the CH component is assumed to have either an intrinsic Gaussian conditional auto-regressive (CAR) prior distribution or a fully specified Multivariate Normal prior distribution [2, 3, 25, 27].

3.1.2 Conditional Auto-regressive (CAR) Models

CAR models provide a tool for detecting and identifying regions where disease risks are clustered. The specification of CAR models provide a framework for borrowing strength between neighboring regions in such a way that, regions that share boundaries are likely to have similar risks and regions that are distant apart are likely to show variability with regard to risk. Waldo Tobler's [28] noted that "*everything is related to everything else but near-by things are more related than distant things*". CAR models were rarely used to detect and cluster risk until the 1990s [13, 17]. The models enable the influence of disease risk in neighboring regions to be modeled and estimated [2, 3, 29]. Distances or boundaries between the regions are used to determine neighborhood properties in the CAR models [18, 20, 30].

Let $\Omega = \{1, 2, \ldots, n\}$ denote the study area and $N_i = \{j \in \Omega : i \in j\}$ classifies regions that share boundaries with region i. Let $\phi_i, i \in \Omega$ be a stochastic variable, then the CH structure of ϕ_i follows a normal distribution defined as:

$$\phi_i \mid \phi_{j\neq i} \sim N\left(\sum_{i\neq j} W_{ij}\phi_j, \tau_i^2\right) \tag{2}$$

where W_{ij} is a spatial dependence parameter for quantifying the weight of each observation on the CAR structure ϕ_i, τ_i^2 is the variance of ϕ_i, and ϕ_j is a set of all observation except ϕ_i. The spatial dependence parameter W_{ij} is non-zero if $j \in S$, but set to zero if $i = j$, in order to prevent auto-correlation [3, 18, 25, 27, 31]. It can be observed from Model 2 that the ϕ_i depends only on a set of neighbors ϕ_j provided the location j is in the neighborhood N_i of ϕ_i.

Assume that region i has M neighbors and $W_{ij} = \frac{1}{M}$ for each region that is a neighbor but zero elsewhere. The conditional expectation of ϕ_i is given by:

$$E\left[\phi_i \mid \phi_{j\neq i}\right] = \mu_i + \sum_{j\in N_i} \Phi_{ij}\left[\phi_j - \mu_i\right]$$

and the conditional variance is:

$$var\left(\phi_i \mid \phi_j\right) = \tau_i^2.$$

The Gaussian processes are defined by mean and covariance functions [**?**]. Thus, the mean and variance-covariance functions are required to specify the CAR model. It follows that the conditional probability distribution of the CAR structure ϕ_i is defined as [25, 27, 28]:

$$f\left(\phi_i \mid \phi_{j\neq i} \in \Omega\right) = \sqrt{\frac{1}{2\pi\tau_i^2}} \exp\left\{-\frac{\left[(\phi_i - \mu_i) - \rho\sum_{j\in N_t} W_{ij}\left(\phi_j - \mu_j\right)\right]^2}{2\tau_i^2}\right\}, \tag{3}$$

where $\mu_i \in R, \tau_i^2 \in R^+, |\rho| < 1, W_{ij} \in R, W_{ij} = W_{ji}, W_{ii} = 0$. The CAR conditional probability distribution function can be written as [3, 25, 27]:

$$f\left(\phi_i \mid \phi_{j\neq i}\right) = \frac{1}{(2\pi)^{n/2} det\left(\mathbf{B}^{-1}\Sigma\right)^{1/2}} \times \exp\left[-\frac{(\boldsymbol{\phi}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}\mathbf{B}(\boldsymbol{\phi}-\boldsymbol{\mu})}{2}\right],$$

where $\boldsymbol{\mu} \in R^n$ is an $n-$dimensional vector with components $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'$, $\Sigma \in R^{+(n\times n)}$ is a symmetric diagonal matrix with components:

$$\Sigma = diag\left(\tau_1^2, \ldots, \tau_n^2\right)$$

and $\mathbf{B} \in R^{n\times n}$ is an invertible matrix defined as:

$$\mathbf{B} = (\mathbf{I} - \rho W) \quad with \quad B_{ij} = \begin{cases} 1 & i = j, \\ -\rho W_{ij} & j \in N_i, \\ 0 & otherwise. \end{cases}$$

The symmetry of Σ implies that the covariance matrix $\Sigma^{-1}\ \mathrm{B} = \mathrm{B}^{-1}\Sigma$ is symmetric such that $W_{(ij)}\tau_j^2 = W_{(ji)}\tau_i^2, i, j \in S$. The probability distribution Function (3) can alternatively be defined as:

$$\phi_i \mid \phi_{j\neq i} \sim N\left[\mu_i + \rho\sum_{j\in N_i} W_{(ij)}\left(\phi_j - \mu_j\right), \tau_i^2\right], \quad (4)$$
$$i \in \Omega \quad and \quad \phi \sim N\left(\mu, \mathrm{B}^{-1}\Sigma\right)$$

It has been proved that the CAR structure ϕ_i follows the Gaussian distribution by showing that Σ is symmetric, see [3, 32] for details.

3.1.3 Parameter estimation: CAR model

Parameters in the CAR model are estimated using Bayesian hierarchical methods. The TB detection data used in the study are counts (whole numbers), therefore, Poisson distribution is assumed for such data. The unknown risk of TB in any region i represented by ϕ_i. The number of cases and population risk in any region i are denoted by y_i and N_i, respectively. The expected number of cases in region i can then be written as:

$$E_i = rN_i,$$

where

$$r = \frac{\sum_{t=1}^{n} y_i}{\sum_{i=1}^{n} N_i}$$

represents the overall risk in the study population. The corresponding likelihood function is defined as:

$$\ell\left(\theta_i\right) = \prod_{i=1}^{n} \frac{\exp\left(-E_i\theta_i\right)\left(E_i\theta_i\right)^{y_t}}{y_i!} = P(\mathbf{y}, \mathbf{E} \mid \boldsymbol{\theta}).$$

Taking natural logarithm of the likelihood function, differentiating with respect to the disease risk θ_i and equating to zero, it can be shown that the maximum likelihood estimator $\hat{\theta}_i$ of θ_i is

$$\widehat{\theta}_i = \frac{y_i}{E_i},$$

which defines the standardized mortality ratio (SMR) in region i. However, using the Bayesian framework, $y_i \sim$ Poisson $(E_i\theta_i)$, where the Poisson mean $\mu_i = E_i\theta_i$, $\theta_i \sim$ Gamma (α, γ) with shape parameter α and scale parameters γ, respectively. However, these formulations do not incorporate covariates in the modeling process. We introduce covariates through a linear predictor, as in the work [2, 3, 20, P. 84]. The distribution of the response variable is specified by the exponent of the linear predictor as $y_i \sim$ Poisson $(E_i \exp(\eta_i))$, where $\mu_i = E_i \exp(\eta_i)$ is the mean of Poisson distribution. Thus, the relative risk of the disease in region i is defined as:

$$\theta_i = \exp\left(\eta_i\right),$$

where $\eta_i = \mathbf{X}'\boldsymbol{\beta} + \phi_i$, and ϕ_i has a CAR structure.

Using the generalized linear model with a log-link function, we have:

$$\log\left(\mu_i\right) = \log\left(E_i\right) + X'\boldsymbol{\beta} + \phi_i.$$

Bayesian models are defined by the posterior distribution of the D parameter estimates, where the posterior is the product of the data likelihood function and the prior distribution(s) of the parameter estimates. Hence, we define the likelihood function as:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\phi}) = \prod_{i=1}^{n} \frac{\left(E_i \exp\left(\eta_i\right)\right)^{y_i} \exp\left(-E_i \exp\left(\eta_i\right)\right)}{y_i!}$$
$$= P(\mathbf{y}, \mathbf{E}, \boldsymbol{\theta} \mid \boldsymbol{\beta}, \boldsymbol{\phi}).$$

The β parameter estimates are assumed to follow the Gaussian distribution defined:

$$P(\beta) = \left(\frac{1}{2\pi}\right)^{P/2}\left(\frac{1}{\tau_\beta}\right)^{P}\exp\left(-\frac{1}{2}\sum_{p=0}^{P}\frac{\beta_p^2}{\tau_\beta^2}\right),$$

and the prior distribution for the CAR random effect is defined by:

$$P(\phi) = \left[\phi_i \mid \phi_{j\neq i}, \tau_\phi^2\right] \sim N\left(\sum_{j\neq i}\frac{w_{ij}}{w_{ij}^{+}}\phi_j, \frac{\tau_\phi^2}{w_{ij}^{+}}\right)$$
$$\sim CAR\left(0, \tau_\phi^2\right)$$

where w_{ij}^{+} is the number of areas which share boundaries with the i^{th} area [6] with:

$$w_{ij} = \begin{cases} 1 & j = i \\ \varphi(i,j) & j \in N_i : \forall i, j \in S, w_{ij} = w_{ji} \\ 0 & otherwise \end{cases}$$

with $\varphi(i, j)$ quantifying the proximity between regions i and j. That is, if $\varphi(i, j) = 1$, then i and j share a common boundary. The posterior distribution is can be expressed as follows:

$$P\left(\beta, \phi, \tau_\beta^2, \tau_\phi^2 \mid \mathbf{y}, \mathbf{E}, \theta\right) \propto P\left(\mathbf{X}, \mathbf{E}, \theta \mid \beta, \phi, \tau_\beta^2, \tau_\phi^2\right)$$
$$\times P(\beta)P(\phi).$$

The hyperprior distribution for the precision parameters τ_ϕ^2 and τ_β^2 are respectively, $\tau_\phi^2 \sim$ Gamma $(0.05, 0.005)$ and $\tau_\beta^2 \sim$ Gamma $(0.5, 0.05)$. The linear regression coefficient distribution is defined by:

$$\beta \sim N\left(0, \tau_\beta^2\right).$$

3.1.4 The Besag, York and Molli'e (BYM) Model
Clayton and Kaldor [33] were first to propose the BYM framework, and later, Besag et. al. developed it further [27]. The BYM (also known as the convolution model) unifies the CH and UH structures into the same model that is capable of explaining clustering and variability of the disease risk. Although various models have been proposed for smooth risks estimation, the model proposed by Besag et. al. (BYM) [27] have been used extensively in literature. The BYM model is expressed as follows:

$$\eta_i = \mu_i + \phi_i + u_i.$$

As indicated in the previous section, the TB cases follow the Poisson distribution, thus, we have $y_i \sim$ Poisson $\left(E_i \exp\left(\eta_i\right)\right)$, where $\mu_i = E_i \exp\left(\eta_i\right)$. The linear link function is $\eta_i = \mathbf{X}'\beta + \phi_i + u_i$. The log relative risk is $\log\left(\theta_i\right) = \eta_i$. Therefore, the relative risk for area i is defined by:

$$\theta_i = \exp\left(\mathbf{X}'\beta + \phi_i + u_i\right).$$

The log log-link function is defined as:

$$\begin{aligned} \log\left(\mu_i\right) &= \log\left(E_i\right) + \left(\mathbf{X}'\boldsymbol{\beta} + \phi_i + u_i\right), \\ &= \log\left(E_i\right) + \log(\theta_i) \\ &= \log(E_i\theta_i), \end{aligned} \tag{5}$$

where $\mathbf{y}$, $\boldsymbol{\beta}, \mathbf{E}$ and $\boldsymbol{\theta}$ are vectors of responses, parameter estimates, the expected number of TB cases and the relative risk of TB, respectively. The u_i is the region-specific random effect quantifying the variability of relative risk of the disease.

3.1.5 Parameter Estimation: BYM
Parameters in the BYM are estimated using the same formulations discussed in Section 3.1.3, however, u_i is required to be a Gaussian prior distribution given by:

$$P(\boldsymbol{u}) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\tau_u}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{u_i^2}{2\tau_u^2}\right).$$

The resulting posterior distribution can be written as follows:

$$\begin{aligned} &P\left(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\phi}, \tau_\beta^2, \tau_u^2, \tau_\phi^2 \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\theta}\right) \\ &\propto P\left(\mathbf{y}, \mathbf{E}, \theta \mid \boldsymbol{\beta}, u, \boldsymbol{\phi}, \tau_\beta^2, \tau_u^2, \tau_\phi^2\right) \times P(\boldsymbol{\beta}) P(\boldsymbol{u}) P(\boldsymbol{\phi}). \end{aligned}$$

The distributions for the hyper-prior precision parameters are as follows: $\tau_u^2 \sim$ Gamma$(0.5, 0.005)$, $\tau_\phi^2 \sim$ Gamma$(0.5, 0.005)$ and $\tau_\beta^2 \sim$ Gamma$(0.5, 0.01)$ respectively. The regression coefficients β follow Gaussian distributions stated as follows;

$$\beta \sim N\left(0, \tau_\beta^2\right).$$

The estimates, τ_u^2 and τ_ϕ^2 are precision-variance estimates for u and ϕ, respectively, and are used to measure the level of variability of risk among the regions and to cluster risk between neighboring regions [3, 29].

In subsequent sections, we will compare the performance of the CAR with the BYM model using the Deviance Information Criterion (DIC) proposed by [34]. The smaller the DIC, the better the model and vice versa. The TB data used in study are collected over time and hence spatial models alone will not be enough to model the space-time pattern of the relative risk of the disease. The spatial models are constrained for identifying heterogeneity and clustering of risk at a single time point. Several methods have been proposed to account for spatial and temporal patterns of disease risks [26, 27, 35–37].

3.2 Space-time models

In this section, space-time models are presented based on three modeling frameworks developed by Knorr-Held et. al. [19, 38], Bernardineli et. al.[26] and Waller et. al.[38]. These models differ with regards to their space-time interactive structures and inclusion of covariates. Consider region i, in year t, that recorded y_{it} TB cases. The cases follow the Poisson distribution, i.e.:

$$\mathrm{y}_{it} \sim Poisson\left(E_{it} \exp\left(\eta_{it}\right)\right),$$

where the unknown relative risk at region i in time t is:

$$\theta_{it} = \exp\left(\eta_{it}\right),$$

and E_{it} is the expected number of TB cases in region i in time t. The expected number of TB cases represents the number of cases expected if the population of region i has statistical behavior comparable to the standard population N_{it}^s. We express the crude rate of TB cases for region i in time t as:

$$r_{it}^s = \frac{y_{it}^s}{N_i^s},$$

and the number of TB cases expected in region i in time t, as:

$$E_{it} = r_{it}^s N_{it} = \frac{y_{it}^s}{N_{it}^s} N_{it},$$

where N_{it} denotes the observed population, y_{it}^s is the TB cases in the standard population. Thus, the overall crude rate of TB cases is given by:

$$\mathrm{r} = \sum_i^n \sum_t^T \frac{y_{it}^s}{N_{it}^3}$$

and the overall number of expected TB cases is defined by:

$$E = \sum_i^N \sum_t^T r_{it}^s N_{it} = \sum_i^N \sum_t^T \frac{y_{it}^s}{N_{it}^s} N_{it}.$$

Our first space-time model formulations is based on the framework developed by Bernardineli et. al.[26], where the linear predictor η_{it} is:

$$\eta_{it} = \mu + \phi_i + u_i + (\varrho + \delta_i) \times t, \tag{6}$$

where $\phi_i + u_i$ follows the BYM specifications [27] (See Section 3.1.4) with spatial structure (ϕ_i) and unstructured random effects (u_i), ϱv_t is the global linear time trend, $v_i\delta_i$ is the interactive term between space and time [19, 26]. The term v_t represents a vector of temporal weights and the intercept μ quantifies the average TB rate in all the 10 regions. Since the risk takes the form $\theta_{it} = \exp(\eta_{it})$ then:

$$\theta_{it} = \exp(\eta_{it}) = \exp[\mu + \phi_i + u_i + (\varrho + \delta_i) \times t].$$

It follows that the Poisson mean is

$$\mu_{it} = E_{it} \exp[\mu + \phi_i + u_i + (\varrho + \delta_i) \times v_t]$$

and logarithm of the mean is given by:

$$\log(\mu_{it}) = \log(E_{it}) + \mu + \phi_i + u_i + (\varrho + \delta_i) \times t.$$

These formulations suggest that each spatial unit has its own time trend with a spatial intercept $(\mu + \phi_i + u_t)$ and a slope $(\varrho + \delta_i)$. This model assumes a linear time trend in each spatial unit. The parameters to be estimated are $\varphi = \{\varrho, \phi, u, \delta\}$ and the hyper-parameters $\psi = \{\tau_\phi, \tau_u, \tau_\delta\}$.

Adjusting for risk factors X_i of TB cases detection, the model 6 can be written as model7. Now the parameters to be estimated are $\varphi = \{\beta, \varrho, \phi, u, \delta\}$ and the hyper-parameters are $\psi = \{\tau_\phi, \tau_u, \tau_\delta\}$.

$$\eta_{it} = \mu + \sum \beta_i X_i + \phi_i + u_i + (\varrho + \delta_i) \times t \tag{7}$$

It is known that if $\delta_i < 0$ the region-specific trend is less steep than the mean trend. On the other hand, $\delta_i > 0$ implies that the region-specific trend is steeper than the mean trend. Further, $\delta_i \sim Normal(0, \tau_\delta)$.

The second space-time model is based on Waller et. al.[**?**] dynamic non-parametric formulation on the linear predictor:

$$\eta_{it} = \mu + \phi_i + u_i + \vartheta_t + \omega_t, \tag{8}$$

where the terms μ, ϕ_i, u_i follow the same formulation as in the first model. ϑ_t and ω_t structures denote the temporally structured and unstructured random effect, respectively. This model assumes a non-parametric time trend. Covariates are incorporated into Model 8 to estimate $\varphi = \{\mu, \beta, \phi, u, \vartheta, \omega\}$ and $\psi = \{\tau_\phi, \tau_u, \tau_\vartheta, \tau_\omega\}$. The model with the covariates can now be written as:

$$\eta_{it} = \mu + \sum \beta_i X_i + \phi_i + u_i + \vartheta_t + \omega_t \tag{9}$$

The ϑ_t quantifies temporal-structure effect and it is modeled using a random walk through a neighboring structure [19] defined as:

$$\begin{cases} \vartheta_t \mid \vartheta_{-t} \sim \mathrm{N}(\vartheta_{t+1}, \tau_\vartheta) & t = 1 \\ \vartheta_t \mid \vartheta_{-t} \sim \mathrm{N}\left(\frac{\vartheta_{t-1}+\vartheta_{t+1}}{2}, \frac{\tau_\vartheta}{2}\right) & t = 2, \ldots, T-1 \\ \vartheta_t \mid \vartheta_{-t} \sim \mathrm{N}(\vartheta_{t-1}, \tau_\vartheta) & t = T. \end{cases}$$

Finally ϕ_t is specified by means of a Gaussian exchangeable prior: $\omega_t \sim N(0, \tau_\omega)$. Finally ϕ_t is specified by means of a Gaussian exchangeable prior: $\omega_t \sim N(0, \tau_\omega)$.

The third space-time Model 10 is an extension of Model 9 that enables a space-time interaction in order to explain the difference in the time trend of TB cases. It is expressed as follows:

$$\eta_{it} = \mu + \phi_i + u_i + \vartheta_t + \omega_t + \pi_{it} \tag{10}$$

In this model, we estimate $\varphi = \{\mu, \phi, u, \vartheta, \omega, \pi\}$ and $\psi = \{\tau_\phi, \tau_u\tau_\vartheta, \tau_\omega\tau_\pi\}$, where π_{it} is interaction between ϕ_i and u_i. The model assumes that there is no interaction between ϕ_i and ϑ_t, hence, $\pi_{it} \sim N(0, \tau_\pi)$. Incorporating covariates into Model 10, yields Model 11:

$$\eta_{it} = \mu + \sum \beta_i X_i + \phi_i + u_i + \vartheta_t + \omega_t + \pi_{it}. \tag{11}$$

Hence, we now have to estimate $\theta = \{\mu, \beta, \phi, u, \vartheta, \omega, \pi\}$ and $\psi = \{\tau_\phi, \tau_u\tau_\vartheta, \tau_\omega, \tau_\pi\}$. For the interaction term π_{it}, it is assumed that there is spatial or temporal structure on the interaction, then $\delta_{it} \sim \mathrm{N}(0, \tau_\delta)$ [38]. In this study, all the precision parameters are assumed to follow the gamma distribution [19].

In Section 4, we assess and compare the performance of the three modeling frameworks (presented in this section) in estimating the relative risk of TB cases in Ghana. We use the Deviance Information Criterion (DIC) proposed by Spiegelhalter [34]. The smaller the DIC, the better the accuracy of the model and vice versa. We report only the results of the most accurate model.

4 Data Analysis and Results

In this section, we analyze the TB cases data described in Section 2, using the hierarchical space and space-time models introduced in Sections 3.1 and 3.2 respectively. The models are implemented in the **R-software** via the Integrated Nested Laplace Approach (INLA) package [19, 22].

Moreover, we performed accuracy experiments for the space-time models using the Deviance Information Criterion (DIC) developed by Spiegelhalter, in order to ascertain the most accurate model for predictive studies. In the discussion, only results obtained from significant predictors are reported and discussed. Further, in the analysis here, a risk value higher than 1 is classified as high risk while risk lower than 1 is classified as low risk. Risk is classified as normal if it has a value of 1.

Furthermore, the space-time models our discussion here, involve the classical parametric framework (7) (presented by [26]), the dynamic non-parametric framework presented in [38] for the linear predictor Equations (9), and Model (10). Model (10) (is an extension of Model (9)) to incorporate interactions between space and time. This enables us to explain the differences in the time trend of the TB cases across the regions. We refer to Equation (7) as **Model I**, Equation (9) as **Model II** and Equation (10) as **Model III**. Results are reported for experiments that involve adjustment and non-adjustment of covariates.

4.0.1 Results: Spatial BYM without covariates adjustment

In this section, we implemented BYM without covariate adjustments (defined as $\eta_i = \mu + \phi_i + u_i.$). The posterior estimates are presented in Table 1. The maps of the posterior mean for the region-specific relative risks $\zeta_i = \exp(\phi_i + u_i)$ presented Figure 6 are used to identify regions with high risk. High risk is visualized by computing $p(\zeta_i > 1 \mid y)$, for details see [19]. The left panel of the Figure 6 shows that five of the ten regions have high risk of TB cases. The risk profile for TB in Ghana are shown in the figure-legends. In the regional map of Ghana, the darker the region the higher the risk and vice versa. It can be observed that Upper East and Upper West Regions have the highest risk (with values in the range **1.8-3.4**), followed by Volta, Western and Central Regions with risk between **1.1-1.8**. The rest of the regions have low risk of TB detection, specifically, Northern and Ashanti Regions have the lowest risk (with values in the range **0.3-0.6**), followed by Greater Accra and Brong Ahafo Regions with values between **0.6-0.9**. Eastern Region has normal risk (with a value in the range **1-1.1**).

The right panel of the Figure 6 shows that Upper East, Upper West, Volta and Central regions have the highest posterior relative risk (**0.8-1**) of TB detection. The regions have similar risks, however, none of them has risk higher than the national risk 1. Moreover, low relative risk (**0-0.2**) of TB detection is observed in the Northern, Brong Ahafo, Ashanti, and Greater Accra regions, followed by Eastern and Western regions that have relative risk values in the range (**0.2-0.8**).

The results in Table 1 confirm the similarity or clustering of risk in the neighboring regions. This is indicated by the low variability captured by the precision of the spatial structure τ_ϕ. We compute the estimate of the posterior marginal variance to capture the amount of variability explained by the spatial structure. We estimated the spatial structure effect empirically using:

$$s_\phi^2 = \frac{\sum_{i=1}^{n} \left(\phi_i - \bar{\phi}\right)^2}{n-1} \tag{12}$$

(where $\bar{v}$ is the average of ϕ) and compare with the posterior marginal variance for the unstructured effect, provided by σ_u^2 :

$$\text{frac}_{\text{ spatial}} = \frac{s_\phi^2}{\left(s_\phi^2 + \sigma_u^2\right)}.$$

The estimated proportion of spatial structure variance is approximately 10%, implying that only 10% of the variability is explained by the spatial structure. It further explains the remaining higher variability captured by the unstructured random effect u_i component of the BYM. The precision τ_u of the unstructured component of the BYM model indicates that risk is heterogeneous among regions. The exponent of the posterior mean μ (overall mean effect) shows that there is approximately a 3-fold increase in TB infections rate across the 10 regions of Ghana. The corresponding 95% credible interval ranges from 2.21 to 4.66.

4.0.2 Results: BYM with covariate adjustments

In this section, we present results obtained from experiments conducted with seven (7) covariate adjustments

of the BYM:

$$\eta_{it} = \mu + \sum_{i=1}^{7} \beta_i X_i + \phi_i + u_i.$$

Among the baseline predictors stated in Section 2.2, the significant predictors for TB cases in Ghana that yield acurate models include: *HIV prevalence, Tuberculosis cure rate, Tuberculosis success rate, proportion of people with knowledge about TB, proportion of those who know that TB is airborne, proportion in high income group and literacy.* Table 2 presents posterior estimates of the overall mean, fixed effects (i.e. $\beta_1, \ldots, \beta_7$) as well as random effects (i.e. τ_u and τ_π) for the unstructured and structured components of the BYM. The maps of the posterior mean for each region's relative risk (i.e. $\zeta_i = \exp(\phi_i + u_i + \sum_{p=1}^{7} \beta p)$) are presented in Figure 7. We can visualize risk by computing $P(\zeta_i > 1 \mid y)$ [19].

It can be observed in the left panel Figure 7 that Upper East, Brong Ahafo and Western Regions have high and similar detection risks ranging from **1.1-1.8.** The adjusted risk, with the covariates adjustment, is less than the unadjusted risk (risk without covariates in Figure 6). Upper East Region is still among the Regions with high risk of TB detection after covariates adjustment. Upper West Region does not belong to the high risk class while Brong Ahafo and Western Regions have moved to the normal risk class after the covariate adjustments. Greater Accra and Central Regions are the second highest in the high risk class, with risk ranging from **1-1.1**. Upper West, Northern, Volta, and Eastern Regions are in the normal risk class, with risk ranging from **0.9-1.** Further, Ashanti Region is in the low risk class (with values in the range **0.6-0.9**) after covariate adjustments. The right panel of the Figure 7, showed that Brong Ahafo and Western Regions have the highest and similar relative risk (**0.8-1**), while Ashanti Regions has the lowest relative risk (**0.0-0.2**). The rest of the Regions have similar relative risks ranging from **0.2-0.8**.

Table 2 presents the posterior estimates of fixed and random effects of the BYM with covariate adjustments. It can be observed that *TB cure rate* increases the risk of TB cases by approximately 8%. This observation implies that as more cases are detected, more cases are cured and hence TB cases will in general decrease over time. This explains why TB success rate leads to 14% reduction of detection. The results also revealed that *knowledge about TB* significantly increases TB detection by approximately 5%. This behavior is expected because, as people become aware of TB, preventive measures are taken (see the results in Section 4.2). *High income* is associated with 5% reduction in TB cases while *literacy* is associated with 12% increase in cases. *High income* increases the use of health facilities and testing for TB, thus, leading to a reduction of TB cases. *HIV prevalence* lead to 55% reduction in cases.

After adjusting the covariates, we observed similarity/clustering of risk between neighboring regions, (see Figure 7) with low variability of risk among the Regions. This observation is captured by the precision τ_ϕ of the spatial structure in Table 2. Heterogeneity of risk across the regions has reduced after the covariates adjustment.

Furthermore, we computed the posterior marginal variance to determine the amount of variability explained by the spatial structure using the formulations in Equation (12). The results showed that the estimated proportion of spatial structure variance is approximately 5%. This implies that only 5% of the variability is explained by the spatial structure. Much of the variability remaining is captured by the unstructured random effect u_i component of the BYM. The precision τ_u of the unstructured component of the BYM indicates that risk is heterogeneous across regions.

The posterior mean of the exponent μ (overall mean effect) gives an indication that there is approximately 9-fold increase in TB infections rate across the 10 regions in Ghana.

4.1 Result: Space-time models without covariates adjustments

Table 3 presents the DIC, mean deviance $\bar{D}$ and effective number of parameters pD components for the three space-time models. The performance indicators show that the classical parametric formulation (see Equation (7) introduced by Bernardinelli et. al. [26] is the most accurate among the three space-time models. Hence, further discussion will include only the results from that model.

The results in Table 4 show that there is about 8% increase in risk of TB detection across the 10 regions of Ghana. However, this increase is statistically not significant at 5% significance level. As observed in the BYM, TB cases do not significantly increase with time. The precision parameter τ_u shows some level of variability in the risk of TB among the regions, while there is clustering of risk between neighboring regions exhibited by the high precision parameter τ_ϕ for the spatial structure. High precision characterized by τ_δ indicates low variability associated with δ_i. This further indicates that there is less interaction between space and time, as well as global trend ϱ and areas-specific trend δ_i. Hence, the area-specific trend δ_i is less remarked than the mean trend.

The left panel of Figure 8 shows the map of spatial trend ζ_i, for the 10 regions and the right panel is the map of the posterior probabilities defined as $p(\zeta_i > 1 \mid y_i)$. The left panel of the Figure 8 shows that there is high risk of TB in the Eastern and Western regions. There is high risk (i.e. in the range **1.1-1.8**) and clustering among the neighboring regions; Volta, Northern, Central, and Greater Accra regions. Upper East and Ashanti regions have the lowest risk (**0.2,0.6**) followed by Upper West and Brong Ahafo (**0.6-0.9**). The results account for the high variability captured by the unstructured component and the low variability captured by the structured component of the area-specific trend. The posterior probabilities in the right panel of Figure 8 indicate low risk (below 1) and relatively low level of associated uncertainty [19]. The time effect is not significant and there is no significant interaction between space and time. This observation accounts for the inacuracy of **Model II** (i.e. Equation (10)) and **Model III** (i.e. Equation (9)) for the TB data.

Next, we will discuss the posterior probabilities of each region by year. Time has no significant effect on the space-time pattern of TB cases as shown in the maps presented in Figure 9 for 2008, 2010, 2011, and 2012 in the top-left panel, top-right panel, bottom-left panel, and bottom-right panel respectively. The risk of TB infection is almost the same across the 10 regions. In the year 2008, there was relatively high risks (in the range **0.8-1**) in the Northern, Volta, Eastern, Western, Central, and Greater Accra regions, while Upper East, Upper West, Brong Ahafo, and Ashanti regions have low relative risk (i.e. between **0-0.2**). The results exhibit clear clustering of risk among neighboring regions associated with low variability or uncertainty. In the year 2010, all the regions had risks in the range **0.2-0.8**. Similar observations can be made in year 2011 and the rest of the years shown in the figure. The results imply that it is sufficient to use only spatial models to estimate the risk of TB across the 10 regions of Ghana. Next, we will adjust the covariates in the space-time models and study their effects in the next section.

4.2 Results: Space-time models with covariate adjustments

It can be observed in Table 5 once more, that **Model I** of the space-time models has the lowest DIC, mean deviance $\bar{D}$ and a high number of effective parameters pD. The indicators show that the classical parametric formulation of [26] (see Equation 7) is still the most accurate model among the three space-time models, for the TB data. Thus, we proceed further in the discussion with results obtained with this model. Table 6 shows negligible risk of TB across the 10 regions of Ghana. Similarly, the TB cases over time is statistically insignificant as observed previously. The precision parameter τ_u indicates very low variability in the risk of TB detection among the regions and much clustering of risk between neighboring regions exhibited by high precision parameter τ_ϕ values for the spatial structure. High precision characterized by τ_δ indicates lower variability associated with δ_i. This further indicates that there is no significant interaction between space and time as well as global trend ϱ and areas-specific trend δ_i. Hence, the area-specific trend δ_i is less remarked than the mean trend.

The results in Table 6 also revealed that *TB success rate* significantly increases TB cases by 11%. Also, *knowledge about TB* significantly reduces TB cases by approximately 2%, while increasing *TB cure rate*, significantly reduces detection by 8%. Awareness that *TB is airborne* increases TB detection by approximately 25%. That is, more people are willing to participate in TB testing to know their status leading to more case detections. We also observed that *HIV prevalence* and *high income* significantly increases TB detection by 27% and approximately 4%, respectively. *Literacy* significantly reduces the risk of TB detection by approximately 14%.

The left panel of the Figure 10 shows the spatial trend ζ_i for the 10 regions and the right panel shows the posterior probabilities defined by $p(\zeta_i > 1 \mid y_i)$. The left panel shows that TB cases risk is higher and clustered in the Volta, Brong Ahafo, Ashanti, Eastern, Western and Greater Accra regions (ranging from **1-1.1**) while there is low risk in the Upper East, Upper West, Northern, and Central regions. Thus, there is high and similarity/clustering of risks among neighboring regions. After covariates adjustments, there is low risk $(0.6, 1)$ in the Upper East, Upper West, Northern and Central regions. These observations account for the low variability captured by both the unstructured and structured components of the area-specific trend.

It can be observed (in the right pane of Figure 10) that all the regions have low relative risks (i.e. in the interval **0.2-0.8**) after adjusting the covariates. Since the time effect is negligible, there is no interaction between space and time. This observation accounts for the inaccuracy of **Model II** (10) and **Model III** (9) for the TB data. The posterior probabilities of the region after covariates adjustments (in the period 2008-2017), showed that the risks of TB across the 10 regions are the same/clustered in the range **0.2-0.8**.

5 Discussion and conclusion

In this discussion, we have modelled the risk of Tuberculosis (TB) across the 10 regions of Ghana using space-time and Bayesian hierarchical space modelling

frameworks. The TB data were obtained from Ghana Health Service (GHS) and National Tuberculosis Programme (NTP) [10, 11]. The models were implemented in the Integrated Laplace Approach (INLA) [19, 22] in **R software**, and used to study the spatial and space-time patterns of TB risk in Ghana. The study also identified some significant predictors of TB risk.

The spatial model used is based on the BYM [27] formulation. The results from this model (without covariates adjustments) showed that hot-spots of TB cases are located in five (5) regions, i.e.; Upper East, Upper West, Volta, Western, and Central regions. Northern, Ashanti, Greater Accra, Brong Ahafo, Eastern and Western regions have low risk of TB detection. Another notable result is the clustering of risk between neighboring regions (i.e. nearby regions have similar risk). The results also revealed that the unstructured component of BYM (that explains variability of risk among the regions) is significant because the spatial structure only explains a small proportion of risk variability among regions. The results also show that after covariates adjustments, the number of high risk regions reduced from five (5) to three (3) (i.e. Upper East, Brong Ahafo and Western regions). The posterior probabilities in the BYM (with and without covariates adjustments) showed that there is clustering of risk between regions.

Further, our study also revealed that *TB cure rate, TB success rate, knowledge about TB, awareness that TB is airborne, HIV prevalence, percentage of literacy, high income* are important predictors of TB detection across the 10 regions of Ghana. Heterogeneity/variability of the risk reduced across the regions after covariate adjustments. The reduction in heterogeneity is due to low variance of the unstructured component and clustering due to low variability of the spatial or structured component of the BYM. Clustering of risk is evident from Figure 7, where almost all the regions have similar risk.

Furthermore, the study showed that the classical parametric formulation (i.e. Equation (7) called **Model I**) is the most accurate space-time model for the TB data. This model yeilds the lowest DIC, lowest mean deviation and highest effective number of parameters with or without covariate adjustments. Hence, it was selected for further experiments. Results from this model show that the risk of TB does not significantly increase over time. There is some level of heterogeneity in risk over time indicated by the precision of the unstructured component. There is relatively high level of clustering among neighboring regions as well. The results shows that there is less interaction of risk between space and time, as well as global trend and area-specific trend. Hence, the area-specific trend is less remarked than the mean trend.

Clustering of risk is evident per the relative risk profile in Figure 8. The space-time model classifies Eastern, Western, Volta, Northern, Central and Greater Accra regions as the hot-spots of the disease over time. Three of the regions (i.e. Volta, Western and Central regions) are classified as high risk regions, by the BYM without covariate adjustments and the Model I without covariate adjustments. The posterior probability in Figure 8 clearly shows clustering of risk and low level of associated uncertainty. The posterior probabilities over the study period are shown in Figure 8. The figures show that the risk of TB does not change over time.

Moreover, after covariate adjustments, statistical inferences remained unchanged and the classical parametric formulation (in Equation 7) remains the most accurate model for the TB data. The posterior summary statistics in Table 6 showed negligible risk of TB across the 10 regions. Precisions of both the unstructured and structured components indicate clustering of risk among the regions. Therefore, all the regions exhibit similar risk. There is no significant space-time interaction due to low variability captured by τ_δ. The results identify the risk factors under the BYM as significant predictors of TB detection.

Therefore, our study has characterized the spatio-temporal pattern of Tuberculosis (TB) in Ghana, using hierarchical space-time models. The key findings include the identification of hotspots, significant baseline predictors, heterogeneity/clustering of risk across regions and insignificant dependance of TB risk on time.

Acknowledgments
Thank you to the Ghana Health Service and National Tuberculosis Program for making data available for the study.

Funding
This study receives no funding.

Availability of data and materials
The data can be found in: `https://africaopendata.org/dataset/the-health-sector-in-ghana-facts-and-figures-2018/resource/0bcf9b54-3e35-4543-95cd-fd4de953edff`, `https://ghanahealthservice.org/downloads/Monitoring_and_evaluation_plan_for_NTP_0.pdf` and `https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22GH%22`.

Authors' contributions
All authors carried out the literature review. AI wrote Statistical Methodology and performed statistical analyses. AI drafted the manuscript. EAA and FKB also contributed to the writing and the reviewing of the manuscript and also provided consultation regarding analysis and interpretation of findings. All authors read and approved the final version of the manuscript.

Competing interests
The author declares that he has no competing interests.

Consent to publish
Not applicable.

Ethics approval (and consent to participate)

Not applicable

References

1. World, H.O.: Global tuberculosis report 2020: executive summary. World Health Organization **20**(2), 2–45 (2020)
2. Iddrisu, A.-K., Alhassan, A., Amidu, N., *et al.*: Investigating spatio-temporal pattern of relative risk of tuberculosis in kenya using bayesian hierarchical approaches. Journal of Tuberculosis Research **6**(02), 175 (2018)
3. Iddrisu, A.-K., Amoako, Y.A.: Spatial modeling and mapping of tuberculosis using bayesian hierarchical approaches. Open Journal of Statistics **6**(3), 482–513 (2016)
4. Kuupiel, D., Vezi, P., Bawontuo, V., Osei, E., Mashamba-Thompson, T.P.: Tuberculosis active case-finding interventions and approaches for prisoners in sub-saharan africa: a systematic scoping review. BMC infectious diseases **20**(1), 1–14 (2020)
5. Otiende, V., Achia, T., Mwambi, H.: Bayesian modeling of spatiotemporal patterns of tb-hiv co-infection risk in kenya. BMC infectious diseases **19**(1), 1–13 (2019)
6. Zumla, A., Petersen, E., Nyirenda, T., Chakaya, J.: Tackling the tuberculosis epidemic in sub-saharan africa–unique opportunities arising from the second european developing countries clinical trials partnership (edctp) programme 2015-2024. International Journal of Infectious Diseases **32**, 46–49 (2015)
7. Amo-Adjei, J., Awusabo-Asare, K.: Reflections on tuberculosis diagnosis and treatment outcomes in ghana. Archives of Public Health **71**(1), 1–8 (2013)
8. Osei, E., Oppong, S., Adanfo, D., Doepe, B.A., Owusu, A., Kupour, A.G., Der, J.: Reflecting on tuberculosis case notification and treatment outcomes in the volta region of ghana: a retrospective pool analysis of a multicentre cohort from 2013 to 2017. Global health research and policy **4**(1), 1–13 (2019)
9. Osei, E., Oppong, S., Der, J.: Trends of tuberculosis case detection, mortality and co-infection with hiv in ghana: A retrospective cohort study. Plos one **15**(6), 0234878 (2020)
10. Service, G.H.: The Health Sector in Ghana Facts and Figures 2018 (2021). https://africaopendata.org/dataset/the-health-sector-in-ghana-facts-and-figures-2018/resource/0bcf9b54-3e35-4543-95cd-fd4de953edff
11. Service, G.H.: National Tuberculosis Programme, Monitoring and Evaluation (2021). https://ghanahealthservice.org/downloads/Monitoring_and_evaluation_plan_for_NTP_0.pdf
12. WHO: Treatment of Tuberculosis: Guidelines vol. 20, pp. 2–45 (2010)
13. Ghana: Tuberculosis profile: Ghana (2021). https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22GH%22
14. Aryee, G., Kwarteng, E., Essuman, R., Agyei, A.N., Kudzawu, S., Djagbletey, R., Darkwa, E.O., Forson, A.: Estimating the incidence of tuberculosis cases reported at a tertiary hospital in ghana: a time series model approach. BMC public health **18**(1), 1–8 (2018)
15. Abdul, I.W., Ankamah, S., Iddrisu, A.-K., Danso, E.: Space-time analysis and mapping of prevalence rate of tuberculosis in ghana. Scientific African **7**, 00307 (2020)
16. Aronis, J.M., Ferraro, J.P., Gesteland, P.H., Tsui, F., Ye, Y., Wagner, M.M., Cooper, G.F.: A bayesian approach for detecting a disease that is not being modeled. PloS one **15**(2), 0229658 (2020)
17. Fouarge, E., Monseur, A., Boulanger, B., Annoussamy, M., Seferian, A.M., De Lucia, S., Lilien, C., Thielemans, L., Paradis, K., Cowling, B.S., *et al.*: Hierarchical bayesian modelling of disease progression to inform clinical trial design in centronuclear myopathy. Orphanet Journal of Rare Diseases **16**(1), 1–11 (2021)
18. Lawson, A., Lee, D.: Bayesian disease mapping for public health. In: Handbook of Statistics vol. 36, pp. 443–481. Elsevier, ??? (2017)
19. Blangiardo, M., Cameletti, M., Baio, G., Rue, H.: Spatial and spatio-temporal models with r-inla. Spatial and spatio-temporal epidemiology **4**, 33–49 (2013)
20. Lawson, A.B.: Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. CRC press, ??? (2018)
21. Lindgren, F., Rue, H., *et al.*: Bayesian spatial modelling with r-inla. Journal of Statistical Software **63**(19), 1–25 (2015)
22. Schrödle, B., Held, L.: A primer on disease mapping and ecological regression using inla. Computational statistics **26**(2), 241–258 (2011)
23. Rue, H., Martino, S., Chopin, N.: Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. Journal of the royal statistical society: Series b (statistical methodology) **71**(2), 319–392 (2009)
24. Wolpert, R.L., Ickstadt, K.: Poisson/gamma random field models for spatial statistics. Biometrika **85**(2), 251–267 (1998)
25. Best, N., Richardson, S., Thomson, A.: A comparison of bayesian spatial models for disease mapping. Statistical methods in medical research **14**(1), 35–59 (2005)
26. Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M.: Bayesian analysis of space—time variation in disease risk. Statistics in medicine **14**(21-22), 2433–2443 (1995)
27. Besag, J., York, J., Mollié, A.: Bayesian image restoration, with two applications in spatial statistics. Annals of the institute of statistical mathematics **43**(1), 1–20 (1991)
28. Miller, H.J.: Tobler's first law and spatial analysis. Annals of the Association of American Geographers **94**(2), 284–289 (2004)
29. Lawson, A.B., Browne, W.J., Rodeiro, C.L.V.: Disease Mapping with WinBUGS and MLwiN vol. 11. Wiley, ??? (2003)
30. Kyung, M., Ghosh, S.K.: Bayesian inference for directional conditionally autoregressive models. Bayesian Analysis **4**(4), 675–706 (2009)
31. Ntzoufras, I.: Bayesian Modeling Using WinBUGS vol. 698. John Wiley and Sons, ??? (2011)
32. Mariella, L., Tarantino, M.: Spatial temporal conditional auto-regressive model: A new autoregressive matrix. Australian Journal of Statistics **39**(3), 223 (2010)
33. Clayton, D., Kaldor, J.: Empirical bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics, 671–681 (1987)
34. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A.: Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64**(4), 583–639 (2002)
35. Knorr-Held, L., Besag, J., *et al.*: Modelling risk from a disease in time and space. Statistics in medicine **17**(18), 2045–2060 (1998)
36. Roza, D.L.d., Caccia-Bava, M.d.C.G., Martinez, E.Z., *et al.*: 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 (2012)
37. Waller, L.A., Carlin, B.P., Xia, H., Gelfand, A.E.: Hierarchical spatio-temporal mapping of disease rates. Journal of the American Statistical Association **92**(438), 607–617 (1997)
38. Knorr-Held, L., Raßer, G.: Bayesian detection of clusters and discontinuities in disease maps. Biometrics **56**(1), 13–21 (2000)

Figures

Figure 1: Trend of TB cases detection per 100,000 by region for 10 years from 2008 to 2017 in Ghana

Figure 2: Trend of the total TB cases detection per 100,000 for each region from 2008 to 2017 in Ghana

Figure 3: Trend of the total TB cases detection per 100,000 population for each year

Figure 4: Box-and-whisker plot of TB cases detection per 100,000 population for each year

Figure 5: Box-and-whisker plot of TB cases detection per 100,000 population by the 10 old administrative Regions in Ghana

Figure 6: Regional distribution of posterior relative risk of TB in Ghana, using BYM: Relative risk of TB $\zeta_i = \exp(\phi_i + u_i)$ (left panel) and Posterior probabilities $p(\zeta_i > 1 \mid y_i)$ in the model (right panel)

Figure 7: Regional posterior relative risk of TB in Ghana, using BYM with adjusted covariates: Relative risks of TB $\zeta_i = \exp(\phi_i + u_i)$ (left panel) and Posterior probabilities $p(\zeta_i > 1 \mid y_i)$ (right panel)

Figure 8: Spatial and uncertainty of TB risk, using space-time models: Spatial TB risk $\zeta_i = \exp(\phi_i + u_i)$ (left panel) and Uncertainty of spatial effects $p(\zeta_i > 1 \mid y_i)$(right panel)

Figure 9: Spatial pattern of TB cases from 2008 to 2012

Figure 10: Posterior probabilities and relative risk of TB in the ten regions of Ghana, using Model I with covariates adjustments.

Tables

Table 1: Summary statistics: posterior mean, standard deviation (Sd) and 95% credible interval for the fixed and random effects of the BYM.

Fixed effects	Estimate	sd	25%	50%	95% CI
μ	3.022	3.25	2.21	3.02	4.66
Random effects					
τ_u	4.00	1.80	1.45	3.69	8.37
τ_ϕ	78.21	84.74	3.28	51.49	306.00

Table 2: Summary statistics: posterior mean, standard deviation (Sd) and 95% credible interval for the fixed and random effects of the BYM model.

Fixed effects	Estimate	Sd	25%	50%	95%
μ	9085.51	16.54	30.70	9115.48	2653723
β_1	1.081	1.027	1.024	1.081	1.141
β_2	0.855	1.040	0.789	0.855	0.927
β_3	1.046	1.017	1.010	1.046	1.083
β_4	0.897	1.055	0.806	0.897	1.000
β_5	0.450	1.297	0.266	0.450	0.762
β_6	0.946	1.017	0.914	0.946	0.979
β_7	1.116	1.045	1.022	1.116	1.219
Random effects	Estimate	Sd	25%	50%	95%
τ_u	25.80	20.62	3.83	20.41	80.01
τ_ϕ	1834.88	1810.43	121.58	1299.34	6656.65

Table 3: Performance indicators for space-time models

Model	$\bar{D}$	p_D	DIC
Model I	518.6	17.81	536.42
Model II	547.1	10.56	557.62
Model III	546.8	11.00	557.76

Table 4: Summary statistics: posterior mean, standard deviation (Sd) and 95% credible interval for the fixed and random effects of the Model I.

Fixed effects	Estimate	sd	25%	50%	95% CI
μ	1.084	1.192	0.763	1.084	1.539
t	1.004	1.007	0.991	1.004	1.017
Random effects	Estimate	sd	25%	50%	95% CI
τ_u	3.87	1.75	1.39	3.57	8.12
τ_ϕ	1847.48	1842.46	124.71	1301.61	6692.95
τ_δ	934.11	940.92	153.67	655.19	3397.82

Table 5: Performance indicators of space-time models with adjusted covariates

Model	D	p_D	DIC
Model I	520.02	17.65	537.67
Model II	547.46	10.62	558.07
Model III	547.17	11.06	558.23

Table 6: Summary statistics: posterior mean, standard deviation (Sd) and 95% credible interval for the fixed and random effects of the Model I.

Fixed effects	Estimate	sd	25%	50%	95% CI
μ	3.65×10^{-6}	3.50	3.04×10^{-7}	3.67×10^{-6}	4.37×10^{-5}
t	1.006	1.007	0.992	1.005	1.019
β_1	0.920	1.012	0.899	0.920	0.943
β_2	1.114	1.018	1.076	1.114	1.155
β_3	0.978	1.008	0.963	0.978	0.993
β_4	1.245	1.023	1.189	1.245	1.302
β_5	2.273	1.120	1.811	2.277	2.829
β_6	1.038	1.008	1.021	1.038	1.052
β_7	0.858	1.019	0.828	0.858	0.891
Random effects	Estimate	sd	25%	50%	95% CI
τ_u	835.68	1220.69	2.33	337.68	4258.23
τ_ϕ	1272.83	1543.46	24.97	738.52	5456.29
τ_δ	521.70	400.33	114.09	412.56	1581.45

Figures

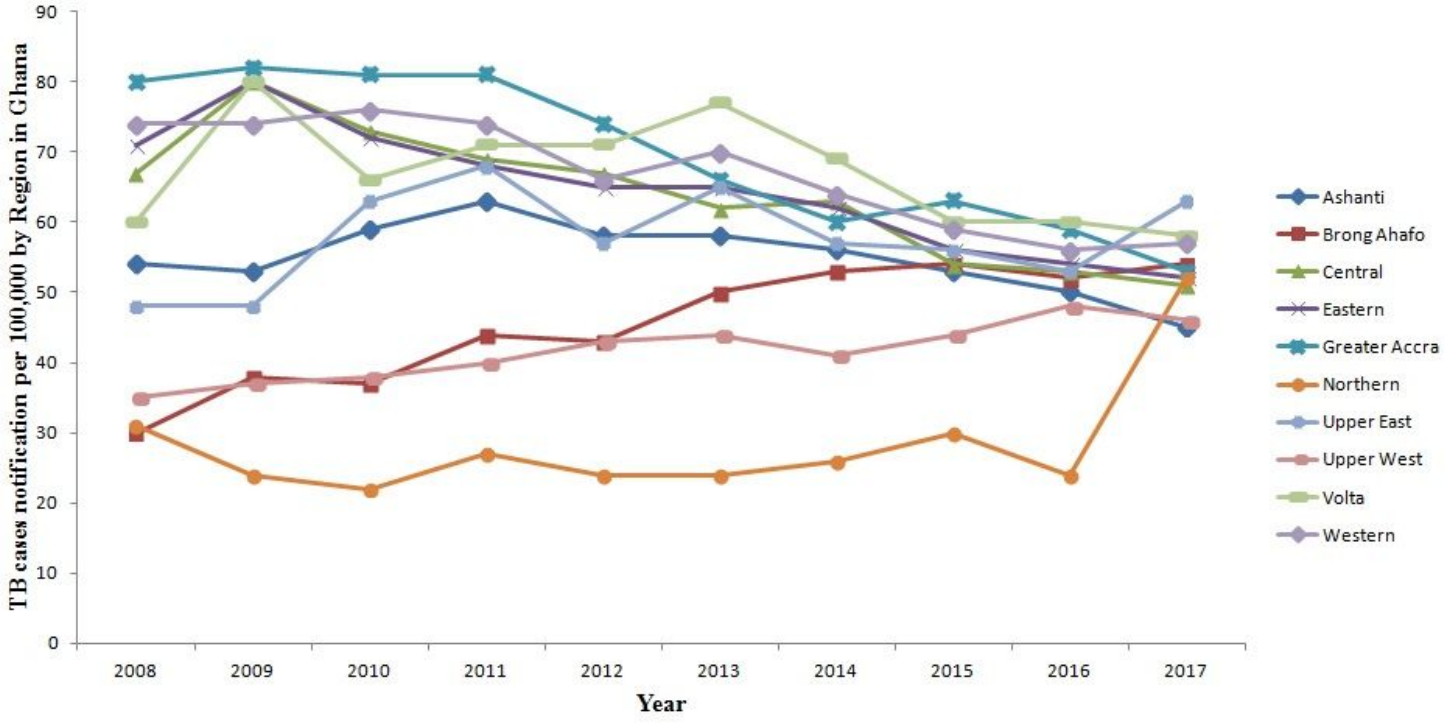

Figure 1

Trend of TB cases detection per 100,000 by region for 10 years from 2008 to 2017 in Ghana

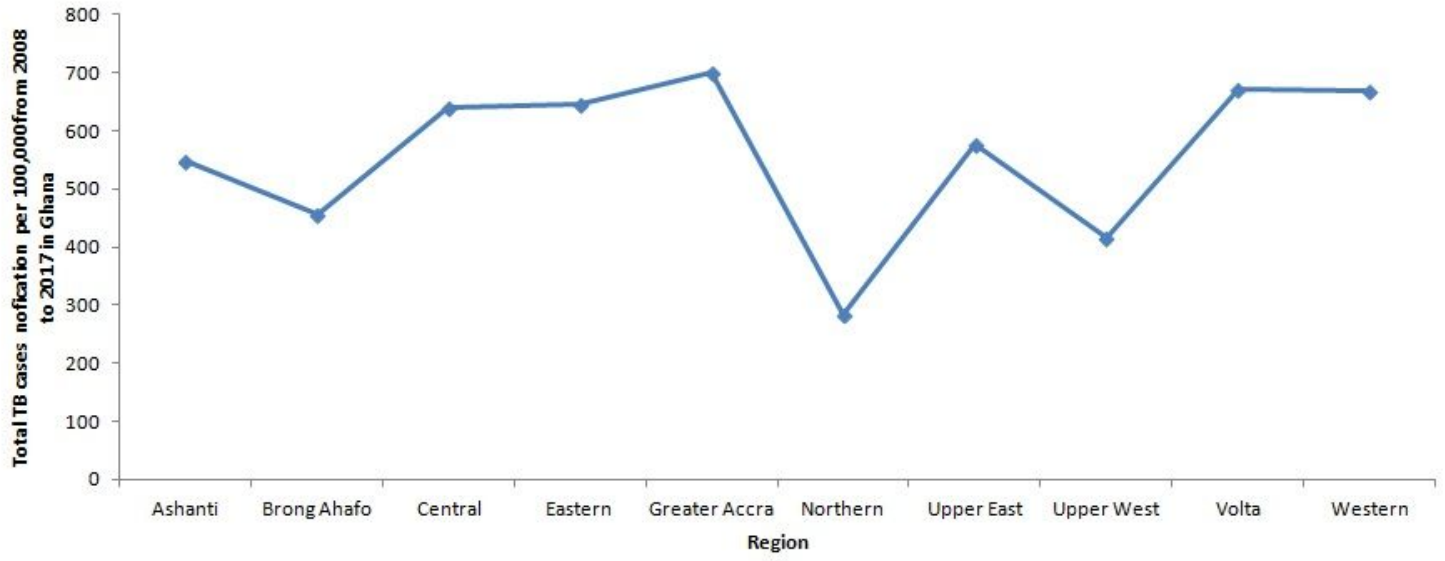

Figure 2

Trend of the total TB cases detection per 100,000 for each region from 2008 to 2017 in Ghana

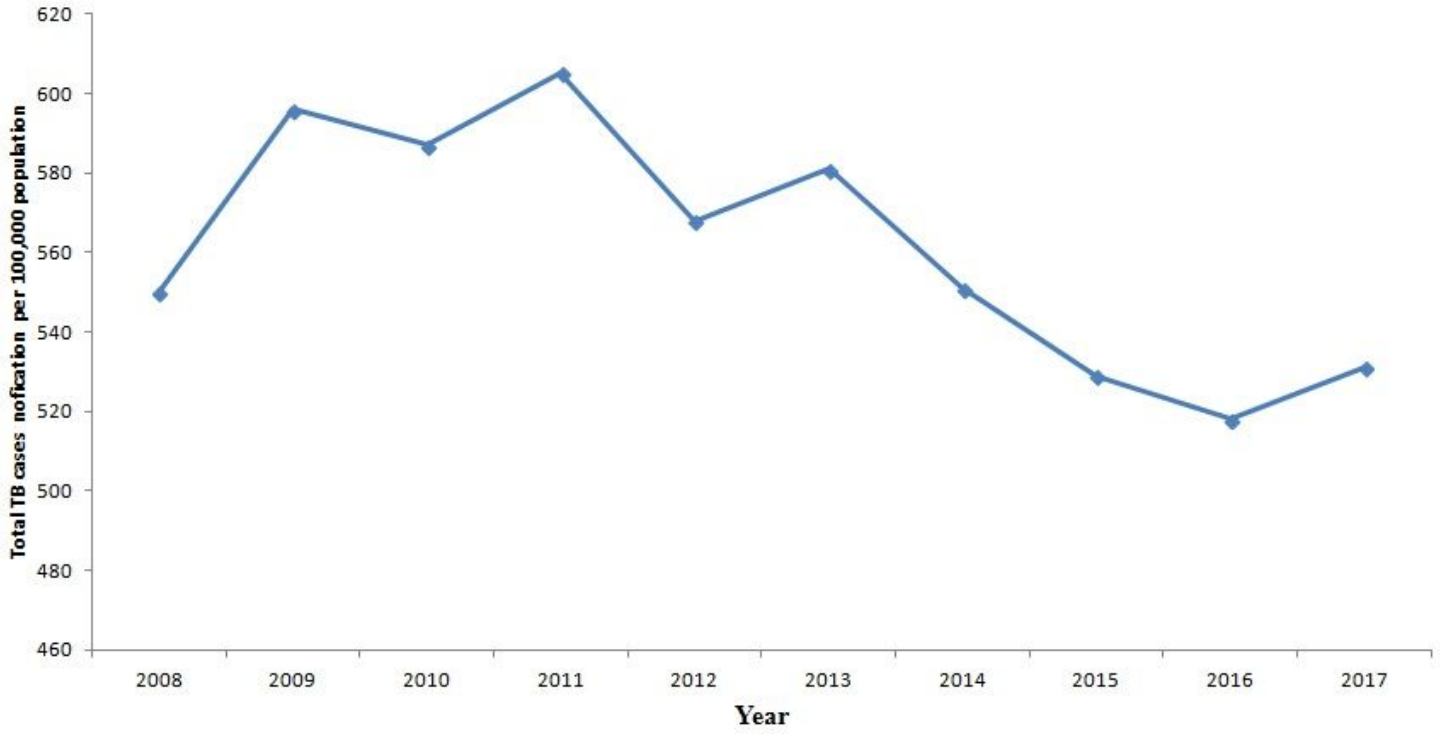

Figure 3

Trend of the total TB cases detection per 100,000 population for each year

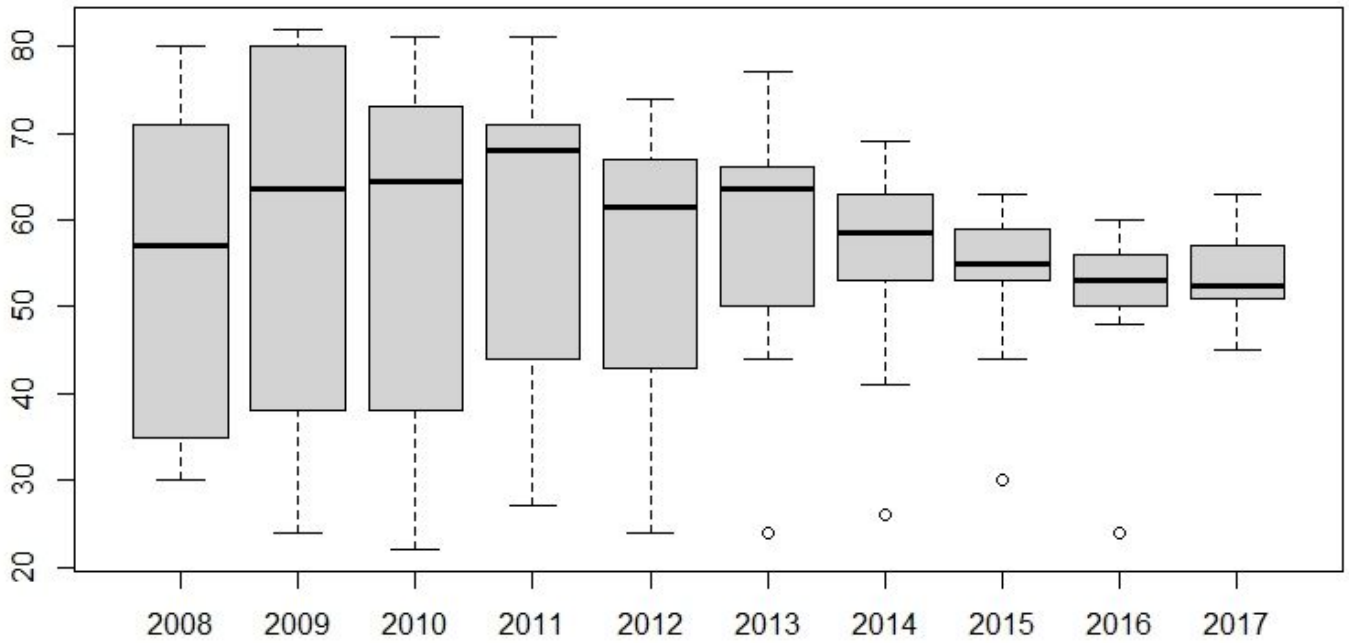

Figure 4

Box-and-whisker plot of TB cases detection per 100,000 population for each year

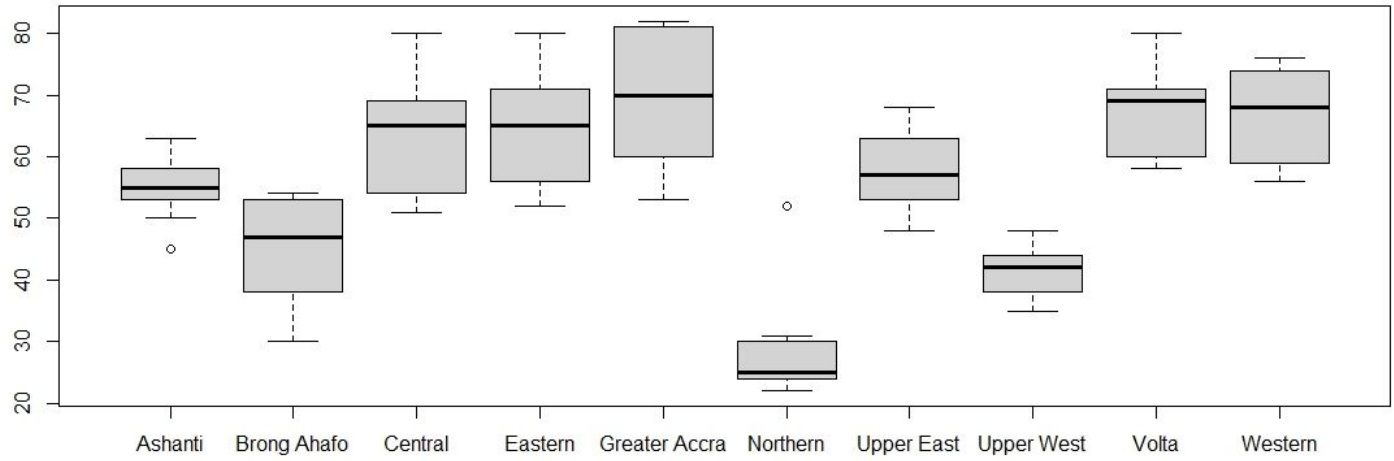

Figure 5

Box-and-whisker plot of TB cases detection per 100,000 population by the 10 old administrative Regions in Ghana

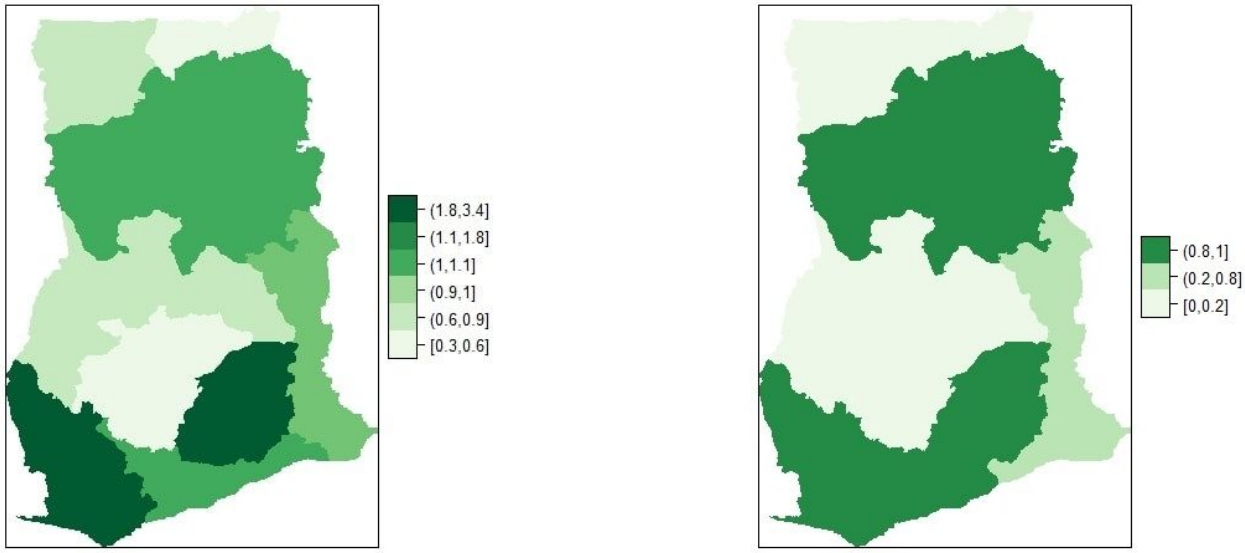

Figure 6

Regional distribution of posterior relative risk of TB in Ghana, using BYM: Relative risk of TB ζi = exp (φi + ui) (left panel) and Posterior probabilities p (ζi > 1 | yi) in the model (right panel)

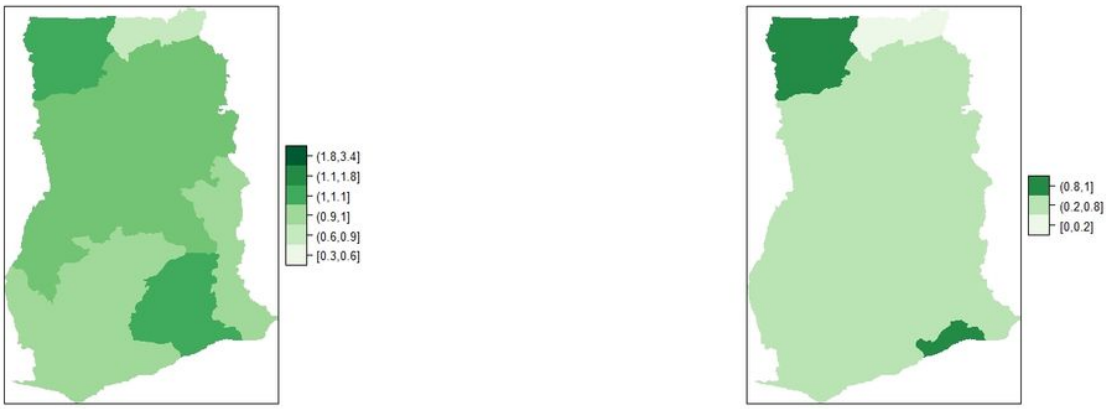

Figure 7

Regional posterior relative risk of TB in Ghana, using BYM with adjusted covariates: Relative risks of TB ζi = exp (φi + ui) (left panel) and Posterior probabilities p (ζi > 1 | yi) (right panel)

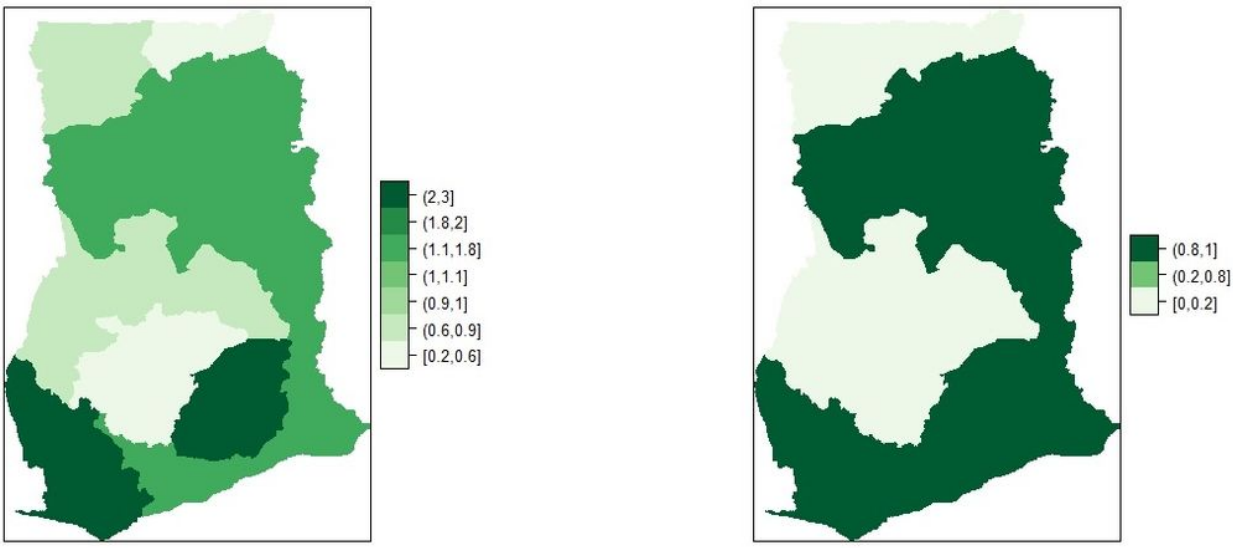

Figure 8

Spatial and uncertainty of TB risk, using space-time models: Spatial TB risk ζi = exp (φi + ui) (left panel) and Uncertainty of spatial effects p (ζi > 1 | yi)(right panel)

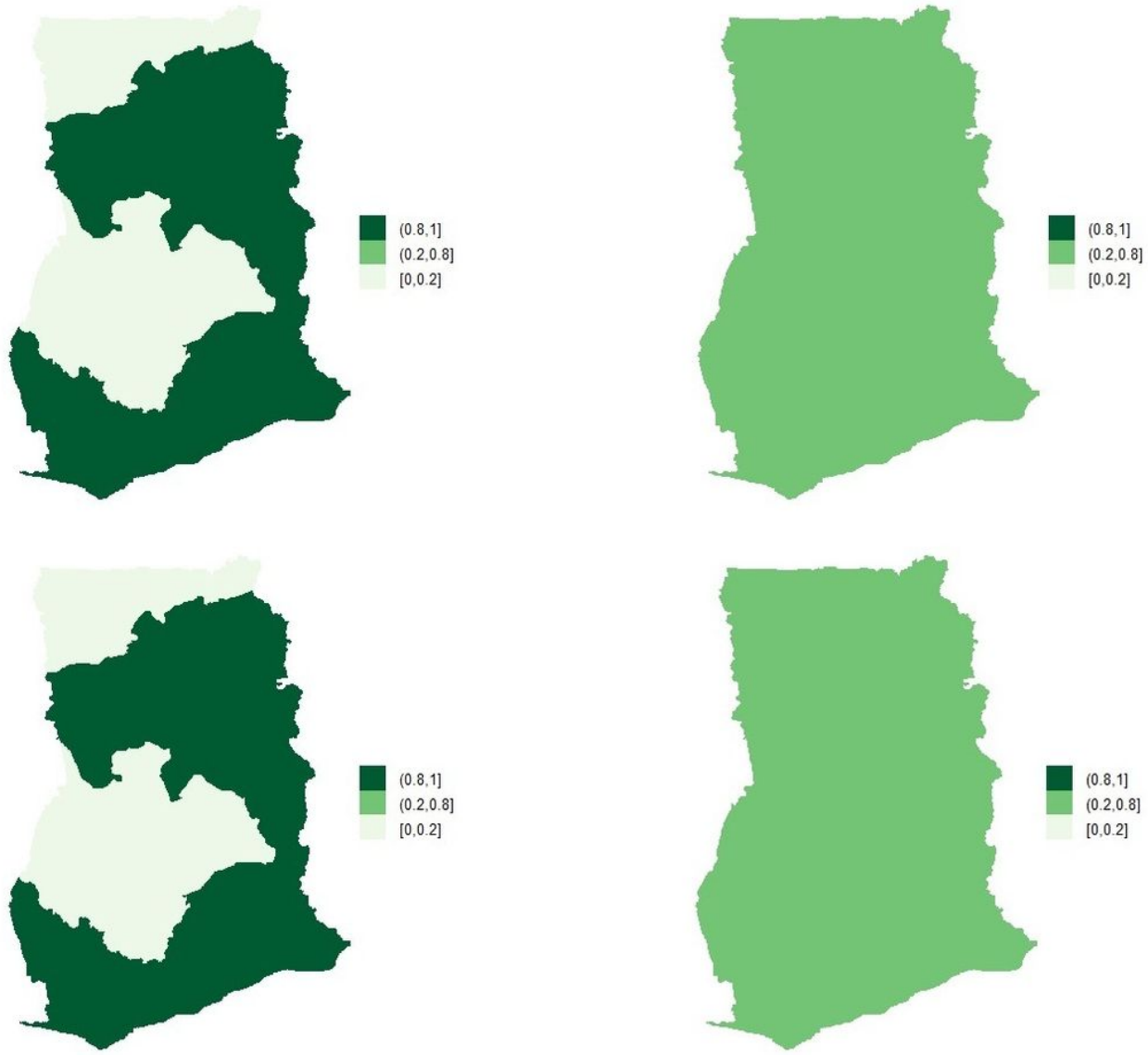

Figure 9

Spatial pattern of TB cases from 2008 to 2012

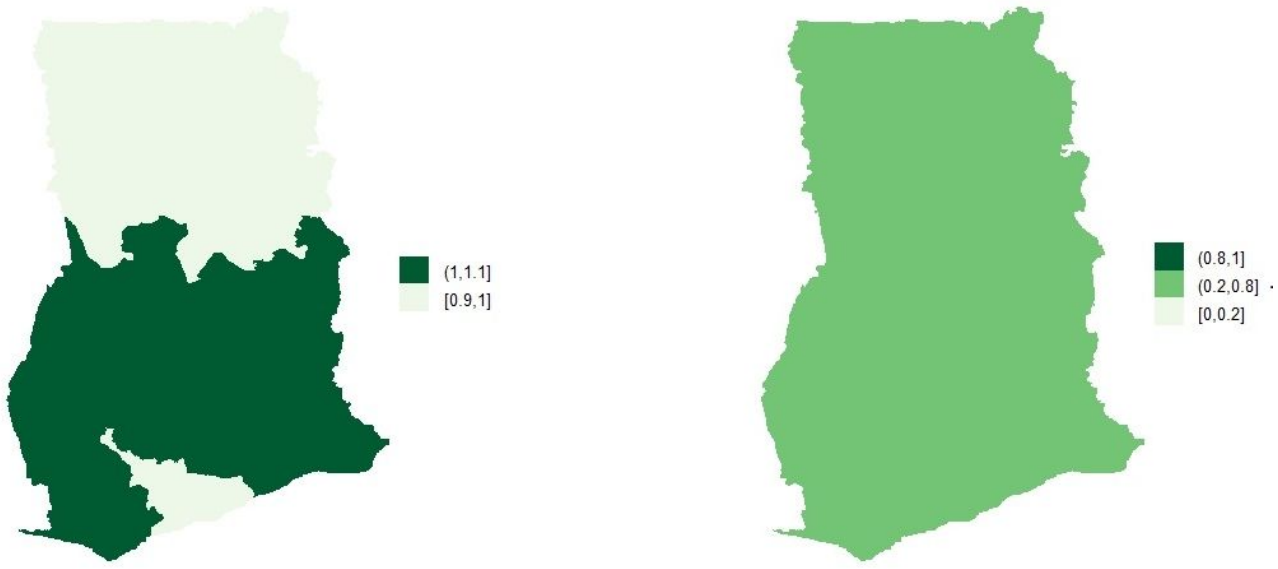

Figure 10

Posterior probabilities and relative risk of TB in the ten regions of Ghana, using Model I with covariates adjustments.