

A Simulation Study of Hierarchical Bayesian Fusion Spatial Small Area Model for Binary Outcome under Spatial Misalignment

Kindie Fentahun Muchie^{1,2*}, Anthony Kibira Wanjoya³, Samuel Musili Mwalili³

¹Institute for Basic Sciences, Technology and Innovation, Pan African University, Nairobi, Kenya

²Department of Epidemiology and Biostatistics, School of Public Health, College of Medicine and Health Sciences, Bahir Dar University, Bahir Dar, Ethiopia

³Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
Email: *muchie.kindie@students.jkuat.ac.ke

How to cite this paper: Muchie, K.F., Wanjoya, A.K. and Mwalili, S.M. (2021) A Simulation Study of Hierarchical Bayesian Fusion Spatial Small Area Model for Binary Outcome under Spatial Misalignment. *Open Journal of Statistics*, 11, 993-1009.
<https://doi.org/10.4236/ojs.2021.116058>

Received: November 4, 2021

Accepted: December 7, 2021

Published: December 10, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Simulation (stochastic) methods are based on obtaining random samples θ^s from the desired distribution $p(\theta)$ and estimating the expectation of any function $h(\theta)$. Simulation methods can be used for high-dimensional distributions, and there are general algorithms which work for a wide variety of models. Markov chain Monte Carlo (MCMC) methods have been important in making Bayesian inference practical for generic hierarchical models in small area estimation. Small area estimation is a method for producing reliable estimates for small areas. Model based Bayesian small area estimation methods are becoming popular for their ability to combine information from several sources as well as taking account of spatial prediction of spatial data. In this study, detailed simulation algorithm is given and the performance of a non-trivial extension of hierarchical Bayesian model for binary data under spatial misalignment is assessed. Both areal level and unit level latent processes were considered in modeling. The process models generated from the predictors were used to construct the basis so as to alleviate the problem of collinearity between the true predictor variables and the spatial random process. The performance of the proposed model was assessed using MCMC simulation studies. The performance was evaluated with respect to root mean square error (RMSE), Mean absolute error (MAE) and coverage probability of corresponding 95% CI of the estimate. The estimates from the proposed model perform better than the direct estimate.

Keywords

Simulation, Small Area Estimation, Hierarchical Bayesian, Spatial Misalignment,

1. Introduction

Simulation (stochastic) methods are based on obtaining random samples of the parameters of interest from the desired distribution and estimating the expectation of any function of the parameters. It can be used for high-dimensional distributions, and there are general algorithms which work for a wide variety of models; where necessary, more efficient computation can be obtained by combining these general ideas with tailored simulation methods, deterministic methods, and distributional approximations. Markov chain Monte Carlo (MCMC) methods have been important in making Bayesian inference practical for generic hierarchical models in small area estimation.

Small Area Estimation (SAE) is a strategy for improving estimation accuracy and delivering reliable parameter estimates for small areas, where a small area refers small sub-population in terms of sample size [1] [2] [3] [4]. To improve accuracy and reliability, a variety of estimators have been developed that combine survey data for the target small areas with data from outside the survey, frequently related to recent censuses and current administrative data.

In small area estimation, statistical models can be utilized in a model assisted approach or model based approach. Even when a model is misspecified, [5] that use working models, in which a model is specified but desirable design based attributes are preserved, are called model assisted estimators.

There are now a variety of model based approaches available, depending on the nature of the measurement (e.g., binary, count or continuous variable) [6] [7] and on the auxiliary data available for the individual units in the population versus those available only at the aggregate level (areal level or unit level) [8] [9].

Unit level estimations are expected to be more precise than areal level estimates on average [10]. Access to auxiliary data at the unit level, on the other hand, may be difficult. As a result, a fusion model that contains both unit and areal level data should be considered to accommodate this. The fusion small area model makes use of census auxiliary data at the area level, which is available in both observation and prediction regions, as well as individual survey variables.

The fusion predictive model links individual survey variables with a hierarchical areal level model, allowing prediction to be carried out to other geographically defined small areas. A fusion small area model, which was successfully used by other authors [11], comprising both unit level survey and areal level auxiliary census data was adopted.

The Bayesian approach to modeling has the ability to combine information from several sources using melding or data fusion [12]. The evaluation of prediction accuracy is an important aspect of SAE [13] although its computation may be complicated. The Bayesian approach solves the problem automatically,

resulting to realizations of the target quantities' from the posterior distribution [13].

A generalized linear model with or without spatial correlation structure was proposed by Ghosh *et al.* [14] using a hierarchical Bayesian framework. In addition, a hierarchical Bayesian model was proposed by Bakar *et al.* [11] for several categorical data under purely spatial setting. Besides, misalignment issues were addressed following this. Furthermore, difficulties with misalignment were resolved as a result of this.

Under the hierarchical Bayesian framework, misaligned data models have recently been presented to solve the issue of making inference for domains other than those for which survey data is available [11] [15] [16]. Following this, Trevisani and Gelfand [17] proposed a model to handle spatial misalignment for count data. For a categorical data under spatial misalignment, Bakar *et al.* [11] had proposed a hierarchical Bayesian model to handle this issue.

Furthermore, Bakar and Jin [18] proposed for estimating a binary outcome variable under purely spatial setting for secondary geography where survey data were not collected. Primary and secondary geographies are spatially misaligned. However, the study by Bakar and Jin [18] considered a real level data.

A fusion model developed by Muchie *et al.* [19] for the binary data has some appealing features. However, it has to be evaluated using simulation studies. As the model has closed forms for all required full conditionals, Gibbs sampler will be used for generating samples.

As a result, the aim of this study was to assess the performance of the hierarchical Bayesian spatial fusion small area model for binary data under spatial misalignment.

2. Materials and Methods

2.1. Model

Here we introduce a spatial hierarchical statistical model for the binary data. Let $Y_i(A_r)$ be a binary response variable for the i^{th} individual in area r , A_r denoting the r^{th} area and also $\mathbf{X}'(A_r)$ be a set of p-covariates in the r^{th} area. Where $i = 1, \dots, N_r$, $r = 1, \dots, n$ and $\sum_{r=1}^n N_r = N$ with $y_i(A_r)$ being a realization of Bernoulli distribution, *i.e.* $y_i(A_r) \sim \text{ber}[P_i(A_r)]$.

Considering y be the realization of a random variable Y that can take values one and zero with probabilities p and $1-p$ respectively. Hence, $y_i(A_r)$ is a Bernoulli response at area A_r and a logit-normal model with probability $p_i(A_r)$ is defined as:

$$\text{logit}\{P_i(A_r)\} = \mu(A_r) + \omega(A_r) + \varepsilon_i(A_r) \quad (1)$$

where $\varepsilon_i(A_r) \sim N(0, \sigma_\varepsilon^2)$ which is the unit level error.

Further,

$$\mu(A_r) = \sum_{l=1}^L \beta_l x_l(A_r) + \zeta(A_r) = \beta X(A_r) + \zeta(A_r) \quad (2)$$

where $x_l(A_r)$ is the l^{th} areal level predictor variable for $l=1,2,\dots,L$, β_l is the corresponding model coefficient, and $\zeta(A_r) \sim N(0, \sigma_\zeta^2)$ is error process; and L is the number of predictors or covariates in the model.

$\omega(A_r)$, is spatially correlated process, modeled explicitly to take the statistical dependency due to spatial proximity. As understanding and defining the spatial process $\omega(A_r) = (\omega(A_1), \dots, \omega(A_n))'$ is often important for constructing the full conditional distribution of ω .

The Moran's I basis function [20] is one of the popular ways to model areal spatial process ω , and consequently, is useful for addressing problems with large spatial data [21]. Hence, for dimension reduction, m eigenvectors of a Moran's I operator matrix are used, where $m < n$. Similarly, dimension reduction approach is used with a set of multiresolution spatial basis function [22] [23]. In multiresolution models [23] m knots are defined over the spatial domain by using reduced rank approximation [24] [25].

Hence, the spatial basis function for the spatial process ω comprises the Moran's I and multiresolution which is written as:

$$\omega(A) = \Phi_{(A)}\eta + \nu \quad (3)$$

where

$$\Phi_{(A)} = M_{(A)} \times R_{(A)} \quad (4)$$

is a combination of eigenvectors $M_{(A)}$ of Moran's I operator matrix $M(A)$ and multiresolution spatial basis $R_{(A)}$. The term

$$\nu \sim N(0, \sigma_\nu^2 \mathbf{I})$$

is an independent and identically distributed error process to capture the remaining random component so as to take the uncertainties arising from $\Phi_{(A)}\eta$.

The process models generated from the predictors (μ) were used to construct the basis so as to alleviate the well-known problem of collinearity [26] between the true predictor variables and the spatial random process. Accordingly, the Moran's I operator matrix is defined as

$$M(A) = (\mathbf{I} - \mu(\mu'\mu)^{-1}\mu')W_{(A)}(\mathbf{I} - \mu(\mu'\mu)^{-1}\mu'), \quad (5)$$

where $\mu = (\mu(A_1), \dots, \mu(A_n))'$, $W_{(A)}$ is $n \times n$ nearest neighbor weight matrix containing 0 s and 1 s, and \mathbf{I} is an $n \times n$ identity matrix.

Spectral decomposition of Moran's I operator matrix $M(A)$ [21] were used to define n eigen vectors, $M_{(A)}$ and multiresolution spatial basis, $R_{(A)}$. A bi-square function was used to define the spatial basis function $R_{(A)}$.

For $j=1, \dots, r_i$ basis functions of the i^{th} resolution, we define the bi-square basis function in d -dimensional Euclidean space \mathbb{R}^d as,

$$S_j^{(i)}(A_r) = \left[1 - \left(\frac{\|A_r - c_j^{(i)}\|}{\phi_i} \right)^2 \right]^2 I(\|A_r - c_j^{(i)}\| < \phi_i); A_r \in \mathbb{R}^d \quad (6)$$

where r_i is the number of basis functions at the i^{th} resolution, $c_j^{(i)}$ is the center of the j^{th} basis function $S_j^{(i)}(\cdot)$ at the i^{th} resolution, $\|\cdot\|$ is the Euclidean norm, ϕ_i is the radius of its spatial support (sometimes called the aperture), and $I(\cdot)$ is an indicator function.

The choice of $\{\phi_i\}$ determines the multiple resolutions, which are used to capture different dependence scales [22] [23] [25]. Some basis functions with centers outside the study domain are included to accommodate boundary effects [27].

When knot locations are equidistant over the domain, the value of ϕ_i is typically some multiple of the minimum distance between knot locations. [22] use a constant range ϕ_i for all knots of a given resolution. To accommodate knot locations that may not be equidistant, we let the ranges of the knots vary. Specifically, we define ϕ_i as a multiple of the minimum distance from knot points to all the other knots. In practice, ϕ_i is specified to be 1.5 times the minimum distance between basis function centers of the same resolution [22].

Another issue that is related to $R_{(A)}$ is the selection of the number and position of knots. A regular grid, or its modification, is often a popular choice for the position of knots [28] [29]. An orthogonal spatial basis $R_{(A)}$ of order $n \times m$ is constructed on the basis of the distances between the centroid of the areas and knot points.

We model η , which is a vector of spatial random effect, by using a Gaussian distribution with zero mean and a lower dimensional covariance matrix Σ , $\eta \sim GP(0, \Sigma)$. Here Σ comprises a smoothing parameter ϕ and $k(\cdot) = \Phi'_{(A)} Q \Phi_{(A)}$ i.e.

$$\Sigma = \phi \times k(\cdot) = \phi \times \Phi'_{(A)} Q \Phi_{(A)}, \quad (7)$$

where Q is $n \times n$ matrix with non-diagonal entries $q_{ij} = 1$ if regions i and j are neighbors and 0 otherwise, and diagonal entries q_{ii} are equal to the number of region i 's neighbors [26]. It can be computed using the neighborhood weight matrix, $W_{(A)}$, as $Q = \text{diag}(W_{(A)} \mathbf{1}) - W_{(A)}$, and $\mathbf{1}$ is a vector of 1 s.

Since Q is singular [30], we use a spectral decomposition of Σ to obtain the full conditional distribution of the spatial process η . Thus, we can write

$$\Sigma = \Psi (\phi \times \Lambda) \Psi', \quad (8)$$

where Ψ is an orthogonal matrix of order $m \times m$ and Λ is a diagonal matrix, see details in [21]. Following [20], in this research we assume that eigen vectors of Σ are known and its eigen values are known up to a multiplicative constant.

2.2. Predictive Distributions

The geographical areas A_k^* ; $k = 1, \dots, n^*$ are those in which prediction is required. Let the area A_k^* have overlapping geographical boundaries with A_j , $j = 1, \dots, n$. We assume that the areal relationships that are learned from A_j are largely maintained for A_k^* as they cover the same region and the same whole popula-

tion. This leads that we can reuse all the parameters with few adjustment of spatial correlation according to the new geographical classification. The fixed effect mean component and the spatial random component were computed separately and joined finally. Specifically, change of support is used to obtain the probability distribution of the process $\mu(A_k^*)$ at a second geographical boundaries.

The change of support for the mean process $\mu(A_k^*)$ in prediction location A_k^* is written as:

$$\mu(A_k^*) = \frac{1}{|A_k^*|} \int_{A_r \subseteq A_k^*} \mu(A_r) dA_r \quad (9)$$

where, $|A^*|$ is the area of the areal unit A^* . The integral in Equation (9) is however difficult to solve, and there are approximation methods available to address this difficulty [31]. However, the approximations are best suited for situations when $A_r \subseteq A_k^*$. This approximation techniques are derived using the concept of the posterior predictive distribution in the Bayesian inference approach.

The posterior predictive distribution is the distribution of possible unobserved values conditional on the observed values. The posterior predictive distribution for unobserved values (μ_k^*) can be found conditional on the observed values (μ_k). To find the predictive posterior distribution of μ_k^* , we considered auxiliary data, X^* is the $n^* \times L$ design matrix, based on second geographical boundaries. Thus, the posterior predictive distribution of $\mu(A_k^*)$ is written as

$$\begin{aligned} \pi(\mu(A_k^*) | \mu) &= \int \pi(\mu(A_k^*) | X^*, \theta, \mu) \times \pi(\theta | X^*, \mu) d\theta \\ &= E_{(\theta | X^*, \mu)} \left[\pi(\mu(A_k^*) | X^*, \theta, \mu) \right]; \end{aligned} \quad (10)$$

where θ contains all the parameters in the model.

On the other hand, for the spatial random effect, new nearest neighbor weight matrix $W_{(A^*)}$ based on $(A^*)_k$'s neighborhood in Moran's I-matrix, and a new orthogonal spatial basis $R_{(A^*)}$ based on the distances between $(A^*)_k$'s centroids and knot points in the space Ω are used. Thus, $\omega(A_k^*)$ is obtained by incorporating $\Phi_{(A^*)}$ in Equation (3). Where, $\Phi^* = M^* \times R^*$ is the basis function that contains W^* neighbourhood weight matrix calculated from at A_k^* . $M_{(A^*)}$ and $R_{(A^*)}$ respectively are the eigen vectors of Moran's I operator matrix and multi-resolution basis based on area A_k^* , i.e. SA2s.

The predictive distribution of $y(A_k^*)$ thus is obtained using MCMC samples and by composition of draws from the posterior distributions and $\pi(\theta | \mu)$.

2.3. Gibbs Sampler Algorithm

The Gibbs sampler, introduced by [32], has been developed by [33]. The Gibbs sampler simulates from multidimensional posterior distributions by iterative sampling from the lower-dimensional full conditional posterior distributions.

The conditional distributions are required for implementation of the model through Markov Chain Monte Carlo (MCMC) simulation. Full conditional densities are derived by abstracting out from the full joint posterior distribution on-

ly those elements including the parameter of interest and treating other components as constants [34]. Accordingly, we obtain the full conditional distribution for the variance parameters and fixed effect coefficients of the model as given below where the detailed derivations are given in a manuscript done by the same authors [19]:

$$\sigma_\varepsilon^2 | \cdot \sim IG \left(\frac{N}{2} + a_\varepsilon, b_\varepsilon + \frac{1}{2} \sum_{i=1}^N \left[\log \left(\frac{P_i(A)}{1 - P_i(A)} \right) - \mu(A) - \omega \right]^2 \right. \\ \left. \times \left[\log \left(\frac{P_i(A)}{1 - P_i(A)} \right) - \mu(A) - \omega \right] \right); \quad (11)$$

$$\sigma_\zeta^2 | \cdot \sim IG \left(\frac{n}{2} + a_\zeta, b_\zeta + \frac{1}{2} [\mu(A) - \beta X(A)]' [\mu(A) - \beta X(A)] \right); \quad (12)$$

$$\sigma_v^2 | \cdot \sim IG \left(\frac{n}{2} + a_v, b_v + \frac{1}{2} [\omega - \phi_{(A)} \eta]' [\omega - \phi_{(A)} \eta] \right); \quad \text{and} \quad (13)$$

$$\phi | \cdot \sim IG \left(\frac{m}{2} + a_\phi, b_\phi + \frac{1}{2} [\eta' \Psi \Lambda^{-1} \Psi' \eta] \right) \quad (14)$$

$$\beta | \cdot \sim N \left(\left(\frac{1}{\sigma_\zeta^2} X X' + \frac{1}{\sigma_\beta^2} I_p \right)^{-1} \left(\frac{1}{\sigma_\zeta^2} X' \mu + \frac{1}{\sigma_\beta^2} \mu_\beta \right), \left(\frac{1}{\sigma_\zeta^2} X X' + \frac{1}{\sigma_\beta^2} I_p \right)^{-1} \right) \quad (15)$$

The Gibbs sampler updates the chain one component at a time, instead of updating the entire vector. In our case, starting from an initial values $(\sigma_\varepsilon^2)^0$, $(\sigma_\zeta^2)^0$, $(\sigma_v^2)^0$, $(\phi)^0$ and $(\beta)^0$, at iteration t , the Gibbs sampler draws:

$$(\sigma_\varepsilon^2)^t \sim \pi \left(\sigma_\varepsilon^2 | (\sigma_\zeta^2)^{t-1}, (\sigma_v^2)^{t-1}, (\phi)^{t-1}, (\beta)^{t-1} \right)$$

$$(\sigma_\zeta^2)^t \sim \pi \left(\sigma_\zeta^2 | (\sigma_\varepsilon^2)^t, (\sigma_v^2)^{t-1}, (\phi)^{t-1}, (\beta)^{t-1} \right)$$

$$(\sigma_v^2)^t \sim \pi \left(\sigma_v^2 | (\sigma_\varepsilon^2)^t, (\sigma_\zeta^2)^t, (\phi)^{t-1}, (\beta)^{t-1} \right)$$

$$(\phi)^t \sim \pi \left(\phi | (\sigma_\varepsilon^2)^t, (\sigma_\zeta^2)^t, (\sigma_v^2)^t, (\beta)^{t-1} \right)$$

$$(\beta)^t \sim \pi \left(\beta | (\sigma_\varepsilon^2)^t, (\sigma_\zeta^2)^t, (\sigma_v^2)^t, (\phi)^t \right)$$

The densities on the right hand sides of the above are called the complete conditional distributions or full conditional distributions. We now provide step by step instructions for running this MCMC algorithm in the list below.

1. Choose initial values for σ_ε^2 , σ_ζ^2 , σ_v^2 , ϕ and β and denote them with $(\sigma_\varepsilon^2)^0$, $(\sigma_\zeta^2)^0$, $(\sigma_v^2)^0$, $(\phi)^0$ and $(\beta)^0$ respectively. Let $t = 0$.
2. Let $t = t + 1$.
3. Update the components of σ_ε^2 , σ_ζ^2 , σ_v^2 , ϕ and β .
 - Let $(\sigma_\varepsilon^2)^t$, $(\sigma_\zeta^2)^t$, $(\sigma_v^2)^t$, $(\phi)^t$ and $(\beta)^t$ equal to a value generated from the full conditionals of σ_ε^2 , σ_ζ^2 , σ_v^2 , ϕ and β respectively.
4. Repeat the steps 2 and 3 until convergence.

Furthermore, the predictive distribution of $y(A_k^*)$ is obtained using MCMC samples and by composition of draws from the posterior distributions and $\pi(\theta | \mu)$. We use the following algorithm to predict $y(A_k^*)$:

1. Draw a sample $\theta^{(t)}$, $t \geq 1$, from the posterior distribution.
2. By creating linear predictors using those parameters drawn in step 1, draw $(\mu(A_k^*))^{(t)}$ from the predictive posterior distributions for $\mu(A_k^*)$ above. That is to replace the above integral by an empirical average of $\pi(\mu(A_k^*) | X^*, \theta, \mu)$ computed using each of those samples in step 1,

$$\left(\pi(\mu(A_k^*) | \mu)\right)^{(t)} \approx \frac{1}{S} \sum_{s=1}^S \pi(\mu(A_k^*) | X^*, \theta^{(s)}, \mu).$$

3. Then draw

$$\left(\log \left\{ \frac{\hat{P}_i(A_k^*)}{1 - \hat{P}_i(A_k^*)} \right\} \right)^{(t)}$$

from the bernouli logit model given above for the i^{th} individual, where $i = 1, \dots, N_k$ for each area A_k^* . Hence the predictive equation can be written in terms of area A_k^* as

$$\mu(A_k^*)^{(t)} + \omega(A_k^*)^{(t)} + \varepsilon_i(A_k^*)^{(t)},$$

where $\omega(A_k^*)$ is obtained by incorporating $\Phi_{(A^*)}$ in Equation (3). $\Phi_{(A^*)}$ is a combination of $M_{(A^*)}$ and $R_{(A^*)}$, which are the eigen vectors of Moran's I operator matrix and multiresolution basis based on area A_k^* , *i.e.* SA2s. The term $\varepsilon_i(A_k^*)^{(t)}$ is obtained from the predictive conditional Gaussian distribution with mean 0 and variance $\widehat{\sigma}_\varepsilon^2$.

4. Finally for each draw (t) we obtain the predictive probabilities in each A_k^* by averaging over the individual predicted probabilities as

$$\left(\hat{P}(A_k^*)\right)^{(t)} = \left(\frac{\sum_{i \in A_k^*} \hat{P}_i(A_k^*)}{N_k} \right)^{(t)}.$$

Steps 1 - 3 in this algorithm are for predicting individual level probabilities. However, interest is in obtaining predictions at the aggregated areal level A_k^* . Hence, in step 4, we aggregate the results by averaging the individual predicted probabilities in each A_k^* for each draw (t). Note that, in the modelling hierarchy, probabilities are aggregated only at the prediction stage.

2.4. Measures of Model Evaluation

The R iterations for the direct and SAE model estimates were evaluated for bias, relative bias, mean squared error (*MSE*), root mean squared error (*RMSE*), percent coverage for a 95% nominal rate, and 95% confidence interval width for direct estimates and 95% credible interval width for the SAE models. Each of these metrics were calculated in two ways: 1) for an individual small area, and 2) average across all individual small areas.

Accordingly the first, *Bias* was calculated as the average difference between the posterior mean of the population total, $\hat{\theta}_{ij}$ of small area i and iteration j , and the truth Y_{Ti} :

$$Bias_i = \frac{1}{R} \sum_{j=1}^R \hat{\theta}_{ij} - Y_{Ti};$$

$$Bias = \frac{1}{n} \sum_{i=1}^n Bias_i$$

The second, Relative bias, RB_i , is defined as the bias relative to the truth for each small area ($RB_i = Bias_i/Y_{Ti}$) and summed across all small areas

$$(RB = \frac{1}{n} \sum_{i=1}^n RB_i).$$

The third, Mean squared error (*MSE*) was calculated as the average squared deviations of the posterior mean of the population total from the true population, and *RMSE* was the square root of these deviations:

$$MSE_i = \frac{1}{R} \sum_{j=1}^R (\hat{\theta}_{ij} - Y_{Ti})^2;$$

$$MSE = \frac{1}{n} \left(\sum_{i=1}^n MSE_i \right)$$

where; Y_{Ti} refers true parameter value for the i^{th} small area, $\hat{\theta}_i$ is the posterior mean of the population total, and R refers the number of iterations. A trade-off between bias and precision (mean squared error, ...) is present when applying a SAE model.

The last one, Percent coverage was defined as the average number of times the 95% SAE credible interval for θ_{ij} , or the direct estimate 95% confidence interval, included truth for each small area. The proportion of times the 95% CI of the estimated summary mean contains the true value.

$$PC_i = \frac{1}{R} \sum_{j=1}^R I(Y_{Ti} \in [95\% \text{ CI of } \hat{\theta}_{ij}]);$$

$$PC = \frac{1}{n} \left(\sum_{i=1}^n PC_i \right)$$

It is desirable to have coverage near 95%. Coverage higher than 95% indicates an inefficient estimator whereas coverage less than 95% indicates an inaccurate estimator.

3. Results and Discussion

3.1. Simulation Design

Proper and weakly informative prior distributions for the model parameters were used. For the fixed effect coefficients (β) we used conjugate normal distribution as the prior distribution with mean zero and a variance of 10 to represent weakly informative and proper prior specification. For the variance and smoothing parameters we consider a conjugate inverse-gamma prior distributions, weakly

informative hyper-parameters for the inverse-gamma prior.

As noted in [35], inverse-gamma prior with very small hyper-parameters does not have any proper limiting posterior distribution. Hence, we consider a proper and weakly informative inverse gamma prior with reasonable values of the shape hyper-parameter $a = 2$; and the rate hyper-parameters $b = 1$ [36] [37], which is popularly suggested in literature [38] [39] [40]. Therefore, for the variance parameters we let the shape and rate hyper-parameters of the inverse gamma prior distribution be 2 and 1, as the inverse gamma (2, 1) distribution is often treated as proper and weakly informative [36] [37].

To get small area estimation at A_r , a total of 400 square grid cells were considered where $N = 3000$ population points were generated. Two predictor variables, $x_{li}(A_r), l = 1, 2$ were generated and correspondingly the binary response, $y_i(A_r)$ (“yes” = 1, “no” = 0) was considered to generate the population values. The probabilities were hence calculated accordingly in each grid cell by identifying the “yes” simulated respondents. The aggregated values of $x_{li}(A_r)$ at each grid cell A_r were treated as census information in the model.

To clarify the insight of the data layout over the domain of the study area we have clearly stated it as follows. Firstly, we took a random sample of 10% [20] [41] of the total population and identify those points inside the square grid cells. Hence, the sample data did not contain all the square grid cells. **Figure 1 & Figure 2** showed the squared grid cells and a representation of the samples in a grid cell together with the population points that were generated in that cell. **Figure 1** showed that the sampled grid cells (shaded in grey colour) (\square) were superimposed,

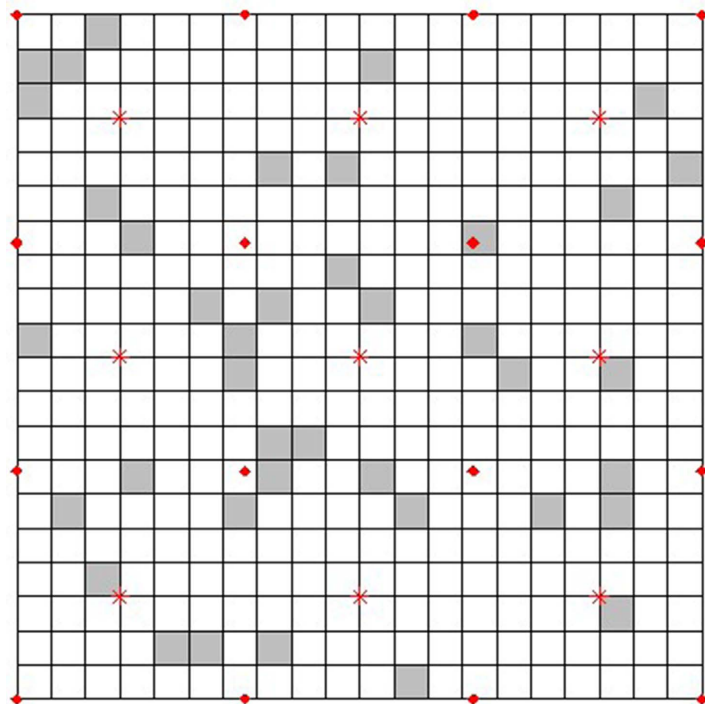


Figure 1. Representation of the population and sample grid cells as well as two resolutions of knot points.

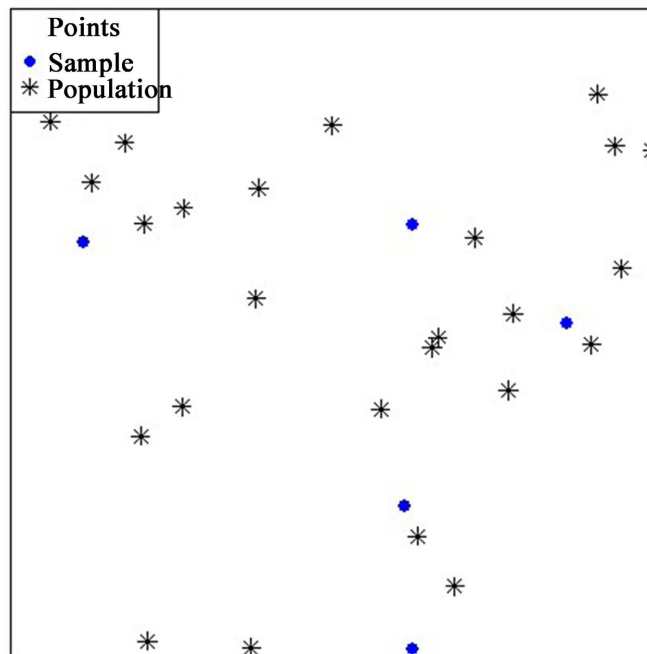


Figure 2. The samples and population points in one squared grid cell.

which is based on randomly selected sample points; two different resolutions of knot points ($*$, \diamond) used for the model proposed were also superimposed.

Secondly, the sample values of the response variable, the aggregated values of auxiliary variables and other spatial components were used to find estimates of hyper-parameters under the SA1.

Thirdly, based on data in A_r , we generate small area estimates for the second geographic areas A_k^* (*i.e.* SA2) having spatial misalignment with the SA1, A_r . The data are generated in a manner similar to that of A_r small area estimation. However, we now aim predictions for $p_i(A_k^*)$, individual level probability at the 100 grid cells, using different knot point resolutions.

3.2. MCMC Output and Diagnosis

Regarding the software implementation, all simulations were performed using the freely available software “R version 4.0.3 (2020-10-10)”. It took 9 hours to generate an MCMC sample of size 50,000 using a machine with Intel(R) Core(TM) i7-4600U CPU @ 2.10 GHz 2.70 GHz and RAM of 8.00 GB (7.90 GB usable).

A gibbs sample of size 50,000 was generated and the first 10,000 samples were discarded as the burn-in period. Further, thinning was applied, every 10th sample selected discarding all others, so as to reduce auto-correlation, to reduce burden on computer memory/storage and to facilitate great deal of post-processing. Hence, 4000 posterior samples were collected after burn-in and thinning. Accordingly, an MCMC sample of size 4000 were used for the entire analysis here under. Accordingly, the trace plot, density plot and auto-correlation plots based on the final 4000 MCMC samples were given below.

Figure 3 showed the trace plots of the MCMC samples for selected model parameters of the model. Each plot illustrated that the chain was sampling from their corresponding stationary distributions. The MCMC chain of variance parameters and the entries of ϕ were mixed well and converged quickly, while the MCMC samples of regression coefficients, β_i , appear to have relatively slow to converge.

The quick convergence of the predictive sample guarantees that the subsequent analyses (such as the averaged predictive means of $\{y_i(A_r)\}$ and $\{p_i(A_r)\}$) are based on appropriate predictive distributions.

Similarly, the kernel density plot (**Figure 4**) also showed there is no serious problem of convergence as the plots are uni-modal. Fast decay of auto-correlation (**Figure 5**) as the lags increase also supplements the above statement regarding convergence.

Henceforth, probabilities for the second geographical areas were computed from posterior predictive distribution. All the following model comparisons were based on these probabilities.

3.3. Model Comparison

The estimates from the model, hierarchical Bayesian fusion spatial small area model, were compared with direct estimate. The RMSE, MAE and 95% probability coverage were considered for comparison purpose. Accordingly, the RMSE

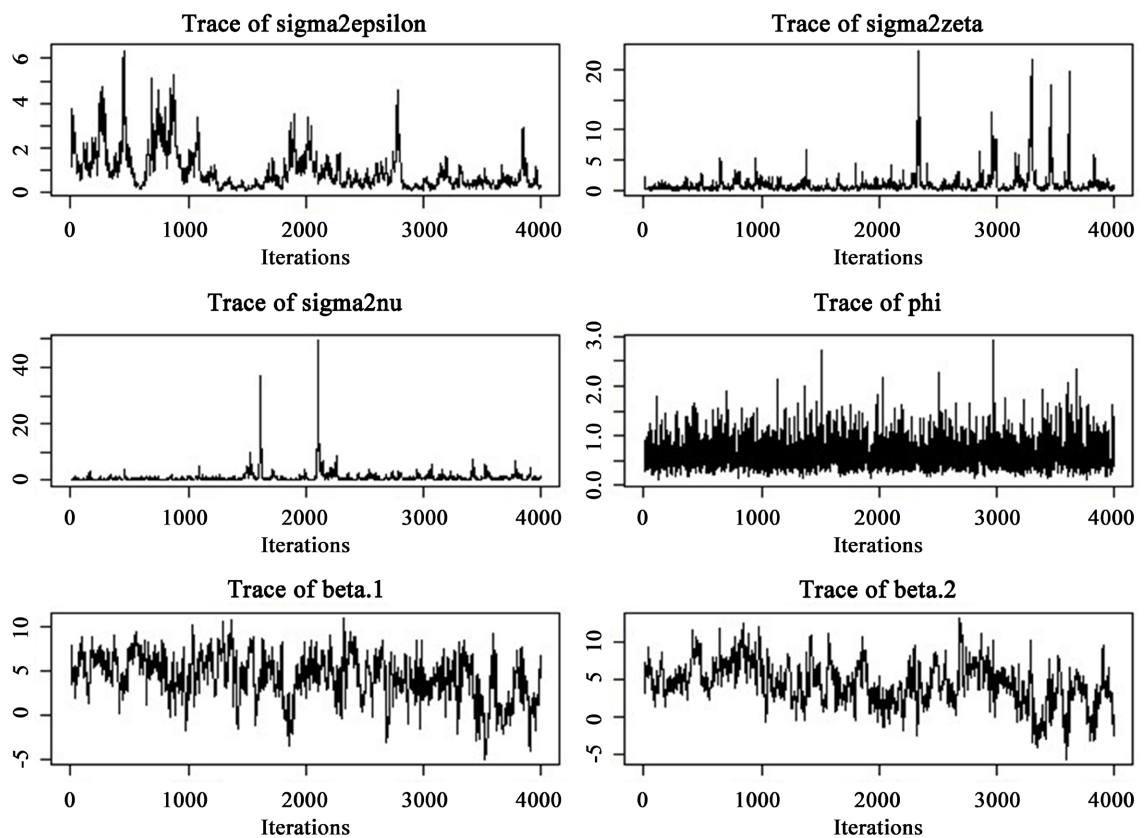


Figure 3. Trace plot of parameters based on MCMC samples.

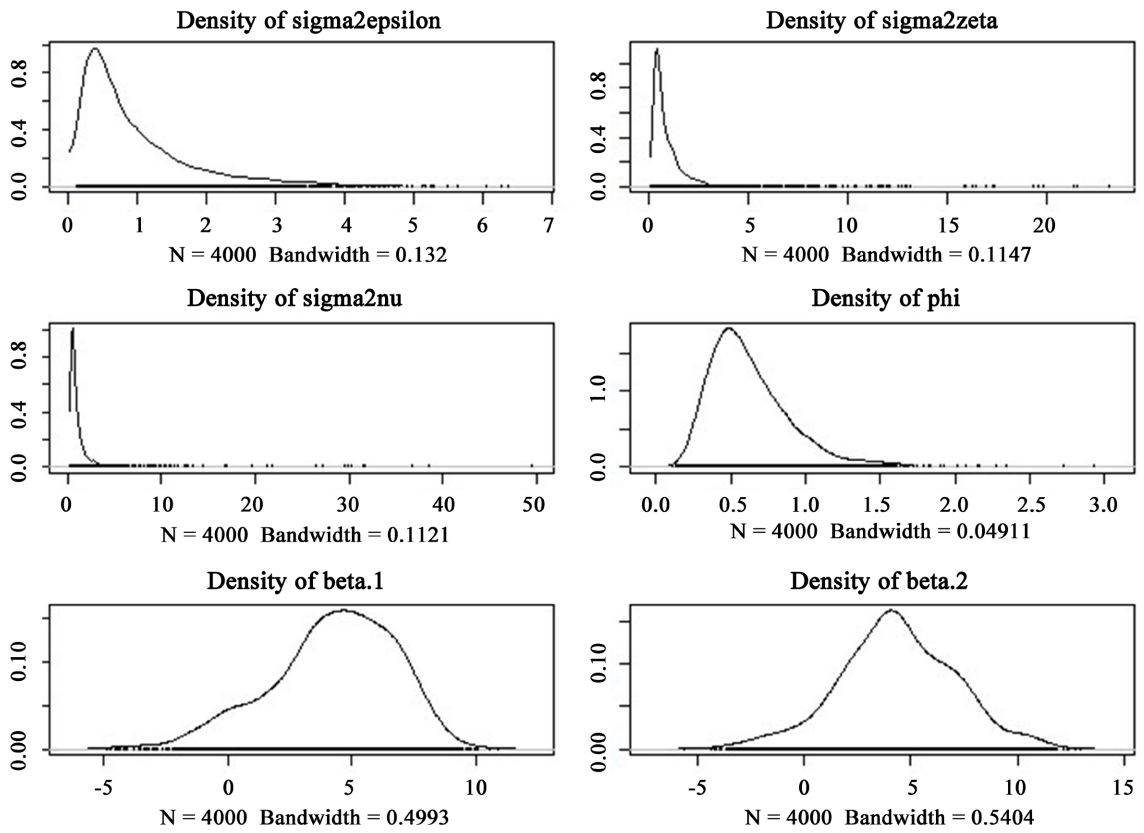


Figure 4. Kernel density plot of parameters based on MCMC samples.

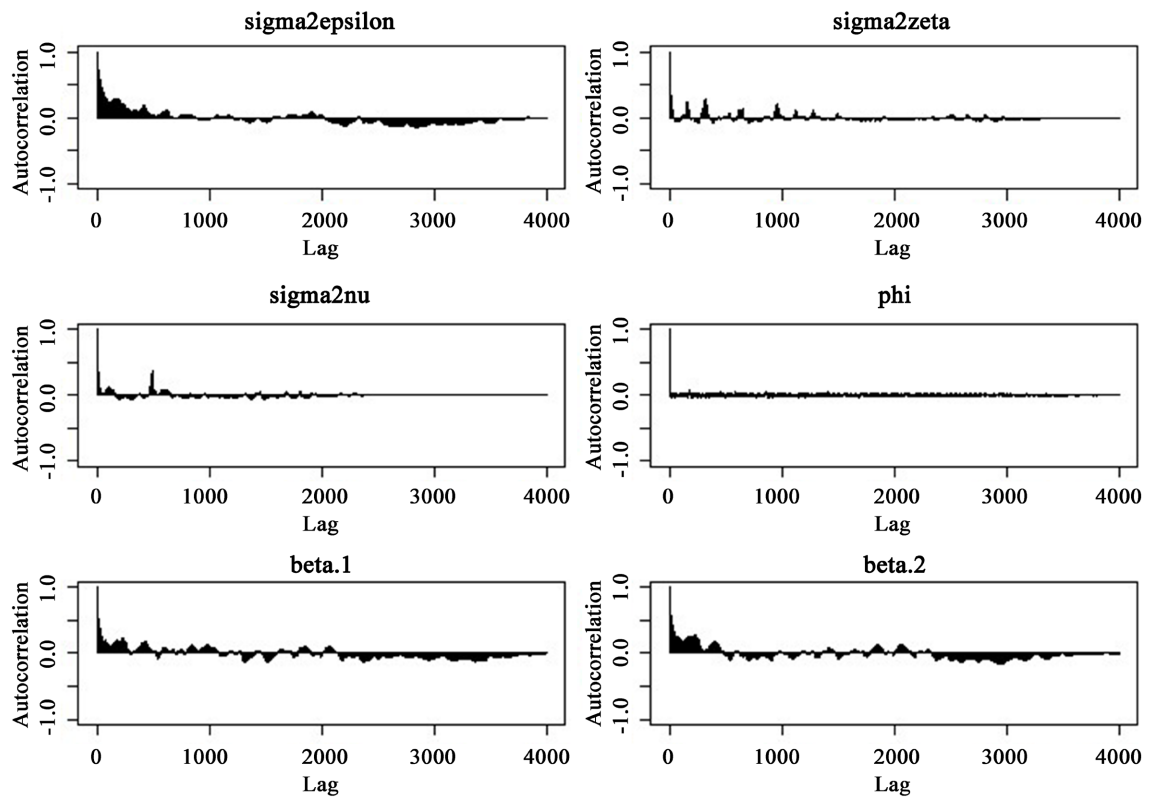


Figure 5. ACF plot of parameters based on MCMC samples.

Table 1. MCMC summary statistics.

Model	RMSE	MAE	Coverage probability
Proposed	0.45	0.43	7.5%
Direct	0.49	0.47	3.8%

and MAE of the estimates from proposed model is less than that of the direct estimate showing the proposed model performs better (**Table 1**). Additionally, the 95% probability coverage for the estimate from the proposed model was greater than that of the direct estimate. This also strengthens the improvement of the proposed model.

4. Concluding Remarks

In this study, we have assessed the performance of the recently developed spatial hierarchical Bayesian fusion spatial small area model for binary data under spatial misalignment. The process models generated from the predictors were used to construct the basis so as to alleviate the well-known problem of collinearity between the true predictor variables and the spatial random process. The performance of the proposed model was assessed using simulation studies. The model proposed has the ability to address the large dimensional spatial problem by reducing dimension through the use of basis function knots. The proposed model performs better than the direct estimate. The developed model can be applied in many broad areas including but not limited to health sciences, public health, agriculture, and economics.

Acknowledgements

We would like to express our appreciation to Pan African University Institute of Basic Sciences, Technology and Innovation for the support.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Rao, J.N. and Molina, I. (2015) Small Area Estimation. John Wiley & Sons, Inc., Hoboken. <https://doi.org/10.1002/9781118735855>
- [2] Datta, G. and Ghosh, M. (2012) Small Area Shrinkage Estimation. *Statistical Science*, **27**, 95-114. <https://doi.org/10.1214/11-STS374>
- [3] Ghosh, M. (1992) Hierarchical and Empirical Bayes Multivariate Estimation. *Institute of Mathematical Statistics Lecture Notes. Monograph Series*, **17**, 151-177. <https://doi.org/10.1214/lnms/1215458844>
- [4] Rao, J.N. and Molina, I. (2015) 1. In: Introduction. 2nd Edition. Small Area Estimation. John Wiley & Sons, Ltd., Hoboken, 1-8.
- [5] Särndal, C.E., Swensson, B. and Wretman, J. (2003) Model Assisted Survey Sam-

- pling. Springer Science & Business Media, Berlin, Heidelberg.
- [6] Fuquene, J. and Betancourt, B. (2011) Heavy Tailed Priors: An Alternative to Non-Informative Priors in the Estimation of Proportions on Small Areas. arXiv: 1107.2724.
 - [7] MacGibbon, B. and Tomberlin, T.J. (1989) Small Area Estimates of Proportions via Empirical Bayes Techniques. *Survey Methodology*, **15**, 237-252.
 - [8] Fay, R.E. and Herriot, R.A. (1979) Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**, 269-277. <https://doi.org/10.1080/01621459.1979.10482505>
 - [9] Rivest, L.P., Verret, F. and Baillargeon, S. (2016) Unit Level Small Area Estimation with Copulas. *Canadian Journal of Statistics*, **44**, 397-415. <https://doi.org/10.1002/cjs.11296>
 - [10] Breidenbach, J., Magnussen, S., Rahlf, J. and Astrup, R. (2018) Unit-Level and Area-Level Small Area Estimation under Heteroscedasticity Using Digital Aerial Photogrammetry Data. *Remote Sensing of Environment*, **212**, 199-211. <https://doi.org/10.1016/j.rse.2018.04.028>
 - [11] Bakar, K.S., Biddle, N., Kokic, P. and Jin, H. (2020) A Bayesian Spatial Categorical Model for Prediction to Overlapping Geographical Areas in Sample Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 535-563. <https://doi.org/10.1111/rssa.12526>
 - [12] Sahu, S.K., Gelfand, A.E. and Holland, D.M. (2010) Fusing Point and Areal Level Space-Time Data with Application to Wet Deposition. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **59**, 77-103. <https://doi.org/10.1111/j.1467-9876.2009.00685.x>
 - [13] Pfeffermann, D. (2013) New Important Developments in Small Area Estimation. *Statistical Science*, **28**, 40-68. <https://doi.org/10.1214/12-STS395>
 - [14] Ghosh, M., Natarajan, K., Stroud, T.W.F. and Carlin, B.P. (1998) Generalized Linear Models for Small-Area Estimation. *Journal of the American Statistical Association*, **93**, 273-282. <https://doi.org/10.1080/01621459.1998.10474108>
 - [15] Gelfand, A.E., Diggle, P., Guttorp, P. and Fuentes, M. (2010) Handbook of Spatial Statistics. CRC Press, Boca Raton. <https://doi.org/10.1201/9781420072884>
 - [16] Ugarte, M.D. (2015) Banerjee, S., Carlin, B.P. and Gelfand, A.E. Hierarchical Modeling and Analysis for Spatial Data. CRC Press/Chapman & Hall. Monographs on Statistics and Applied Probability 135, Boca Raton, Florida, 2015. 562 p. \$ 99.95. ISBN-13: 978-1-4398-1917-3 (Hardcover). *Biometrics*, **17**, 274-277. <https://doi.org/10.1111/biom.12290>
 - [17] Trevisani, M. and Gelfand, A. (2013) Spatial Misalignment Models for Small Area Estimation: A Simulation Study. In: Torelli, N., Pesarin, F. and Bar-Hen, A., Eds., *Advances in Theoretical and Applied Statistics*, Springer, Berlin, Heidelberg, 269-279. https://doi.org/10.1007/978-3-642-35588-2_25
 - [18] Bakar, K.S. and Jin, H. (2020) Areal Prediction of Survey Data Using Bayesian spatial Generalised Linear Models. *Communications in Statistics-Simulation and Computation*, **43**, 2963-2978. <https://doi.org/10.1080/03610918.2018.1530787>
 - [19] Muchie, K.F., Wanjoya, A.K. and Mwalili, S.M. On Hierarchical Bayesian Spatial Fusion Small Area Model for Binary Data under Spatial Misalignment. Under Review.
 - [20] Hughes, J. and Haran, M. (2013) Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **75**, 139-159. <https://doi.org/10.1111/j.1467-9868.2012.01041.x>

- [21] Bradley, J.R., Wikle, C.K. and Holan, S.H. (2016) Bayesian Spatial Change of Support for Count-Valued Survey Data with Application to the American Community Survey. *Journal of the American Statistical Association*, **111**, 472-487. <https://doi.org/10.1080/01621459.2015.1117471>
- [22] Cressie, N. and Johannesson, G. (2008) Fixed Rank Kriging for Very Large Spatial Data Sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**, 209-226. <https://doi.org/10.1111/j.1467-9868.2007.00633.x>
- [23] Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F. and Sain, S. (2015) A Multiresolution Gaussian Process Model for the Analysis of Large Spatial Datasets. *Journal of Computational and Graphical Statistics*, **24**, 579-599. <https://doi.org/10.1080/10618600.2014.914946>
- [24] Cressie, N., Shi, T. and Kang, E.L. (2010) Fixed Rank Filtering for Spatio-Temporal Data. *Journal of Computational and Graphical Statistics*, **19**, 724-745. <https://doi.org/10.1198/jcgs.2010.09051>
- [25] Katzfuss, M. (2017) A Multi-Resolution Approximation for Massive Spatial Datasets. *Journal of the American Statistical Association*, **112**, 201-214. <https://doi.org/10.1080/01621459.2015.1123632>
- [26] Reich, B.J., Hodges, J.S. and Zadnik, V. (2006) Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models. *Biometrics*, **62**, 1197-1206. <https://doi.org/10.1111/j.1541-0420.2006.00617.x>
- [27] Cressie, N. and Kang, E.L. (2010) High-Resolution Digital Soil Mapping: Kriging for Very Large Datasets. In: Viscarra Rossel, R., McBratney, A. and Minasny B., Eds., *Proximal Soil Sensing*, Springer, Dordrecht, 49-63. https://doi.org/10.1007/978-90-481-8859-8_4
- [28] Diggle, P. and Lophaven, S. (2006) Bayesian Geostatistical Design. *Scandinavian Journal of Statistics*, **33**, 53-64. <https://doi.org/10.1111/j.1467-9469.2005.00469.x>
- [29] Sahu, S.K., Bakar, K.S. and Awang, N. (2015) Bayesian Forecasting Using Spatiotemporal Models with Applications to Ozone Concentration Levels in the Eastern United States. In: Dryden, I.L. and Kent, J.T., Eds., *Geometry Driven Statistics*, Vol. 121, John Wiley & Sons, Ltd., Hoboken, 260. <https://doi.org/10.1002/9781118866641.ch13>
- [30] Diggle, P.J. and Ribeiro, P.J. (2007) *Model-Based Geostatistics*. Springer-Verlag, New York. <https://doi.org/10.1007/978-0-387-48536-2>
- [31] Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2014) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, New York. <https://doi.org/10.1201/b17115>
- [32] Geman, S. and Geman, D. (1984) Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**, 721-741. <https://doi.org/10.1109/TPAMI.1984.4767596>
- [33] Gelfand, A.E. and Smith, A.F. (1990) Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 398-409. <https://doi.org/10.1080/01621459.1990.10476213>
- [34] Gilks, W.R. (1995) Full Conditional Distributions. In: Gilks, W.R., Richardson, S. and Spiegelhalter, D., Eds., *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, New York, 75-88.
- [35] Gelman, A. (2006) Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper). *Bayesian Analysis*, **1**, 515-534. <https://doi.org/10.1214/06-BA117A>
- [36] Bakar, K.S. and Kokic, P. (2017) Bayesian Gaussian Models for Point Referenced

-
- Spatial and Spatio-Temporal Data. *Journal of Statistical Research*, **51**, 17-40. <https://doi.org/10.47302/jsr.2017510102>
- [37] Sahu, S.K. and Bakar, K. (2012) A Comparison of Bayesian Models for Daily Ozone Concentration Levels. *Statistical Methodology*, **9**, 144-157. <https://doi.org/10.1016/j.stamet.2011.04.009>
- [38] Gelfand, A.E. and Sahu, S.K. (1999) Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association*, **94**, 247-253. <https://doi.org/10.1080/01621459.1999.10473840>
- [39] Gelman, A., Jakulin, A., Pittau, M.G. and Su, Y.S. (2008) A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. *Annals of Applied Statistics*, **2**, 1360-1383. <https://doi.org/10.1214/08-AOAS191>
- [40] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2013) Bayesian Data Analysis. CRC Press, New York. <https://doi.org/10.1201/b16018>
- [41] Bradley, J.R., Holan, S.H. and Wikle, C.K. (2015) Multivariate Spatio-Temporal Models for High-Dimensional Areal Data with Application to Longitudinal Employer-Household Dynamics. *The Annals of Applied Statistics*, **9**, 1761-1791. <https://doi.org/10.1214/15-AOAS862>