Research Article

On Hierarchical Bayesian Spatial Small Area Model for Binary Data under Spatial Misalignment

Kindie Fentahun Muchie [ID],[1,2] Anthony Kibira Wanjoya,[3] and Samuel Musili Mwalili[3]

[1]*Pan African University Institute for Basic Sciences, Technology and Innovation, Nairobi, Kenya*
[2]*Bahir Dar University, Bahir Dar, Ethiopia*
[3]*Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya*

Correspondence should be addressed to Kindie Fentahun Muchie; muchie.kindie@students.jkuat.ac.ke

Small area models have become popular methods for producing reliable estimates for sub-populations (small geographic areas in this study). Small area modeling may be carried out via model-assisted approaches within the model-based approaches or design-based paradigm. When there are medium or large samples, a model-assisted approach may be reliable. However, when data are scarce, a model-based technique may be required. Model-based Bayesian analysis is popular for its ability to combine information from several sources as well as taking account uncertainties in the analysis and spatial prediction of spatial data. Nevertheless, things become more complex when the geographic boundaries of interest are misaligned. Some authors have addressed the problem of misalignment under hierarchical Bayesian approach. In this study, we developed non-trivial extension of existing hierarchical Bayesian model for a binary outcome variable under spatial misalignment with three contributions. First, the model uses unit-level survey data and area-level auxiliary data to predict the posterior mean proportion spatially at the second geographic area level. Second, the linking model is changed to logit-normal model in the proposed model. Lastly, the mean process was considered to overcome the multicollinearity between the true predictors and the spatial random effect. Sensitivity analysis was also done via simulation.

1. Introduction

Sample surveys are widely used as a cost-effective way to gather information on variables of interest in target populations. Sample survey data are used to obtain precise estimates of parameters for the whole population and for large sub-populations (domains). However, a particular widespread demand from policy makers is for estimates at a finer level of domains.

In sample survey, when the domains were not originally planned, they are usually poorly represented or even not represented at all. These domains are called small areas, for which reliable (precise and accurate) estimates based on only domain-specific sample data cannot be produced [1]. Small areas can be based on geographic boundaries or based on a combination of socio-demographic characteristics like urban/rural, male/female, age group, economic status, and so on. The term "small area" in this study context refers to a sub-population based on geographical classification, i.e., small geographic areas.

In this case, small area estimation (SAE) is a technique for improving estimation accuracy and providing accurate parameter estimates for small areas [2–4]. This has led to the development of a variety of models that combine survey data with data from outside the survey, often related to recent censuses and current administrative data, for the target small areas to improve precision.

Statistical models in small area estimation can be utilised in either a model-assisted or model-based approach. The model-assisted estimators [5] employ working models, in which a model is specified yet desired design-based properties are retained even when the model is misspecified.

When the sample size from the target population is large enough and sampling plan is well, according to a cautious viewpoint [6], model-assisted inference performs satisfactorily. Despite the fact that a survey's sampling strategy and sample size are normally well planned at the population level, due to probability sampling and budget constraints, the sample size in small areas can be small or even zero. In this case, model-assisted estimators for small area parameters cannot achieve desired accuracy and thus become inaccurate, necessitating the use of a model-based approach.

There now exist a wide range of model-based approaches depending on the nature of the measurement (e.g., binary, count, or continuous) and on the availability of auxiliary data (area level or unit level).

On average, unit-level estimates are expected to be more precise than area-level estimates [7]. When unit-level data are available, we can use a model for more in-depth analysis. However, access to unit-level auxiliary data might be challenging. Hence, a fusion model incorporating both unit and area-level data can be considered to accommodate this challenge. The fusion small area model exploits the area-level auxiliary data and individual-level survey data, which are effectively applied by other authors under the Bayesian approach [8].

Bayesian approach of modeling has the ability to combine information from several sources using melding or data fusion [9]. The approach also simplifies computation of measures of accuracy in the SAE, which produces realizations of the posterior distribution of the target quantities [10]. Under the hierarchical Bayesian framework, Ghosh et al. [11] proposed a generalized linear model with or without spatial correlation structure. Furthermore, Bakar et al. [8] proposed a hierarchical Bayesian spatial model for several categorical data under spatial misalignment.

Misaligned data models have recently been presented using the hierarchical Bayesian framework to solve the challenge of making inferences for small areas that differ from those for which survey data are available [8, 12, 13]. Following this, Trevisani et al. [14] proposed a model to handle spatial misalignment for count data. For several categorical variables under spatial misalignment, Bakar et al. [8] proposed a hierarchicthe posterior mean proportional Bayesian model to handle this issue.

Furthermore, Bakar et al. [15] proposed to estimate a binary outcome variable under purely spatial setting for secondary geography where survey data were not collected, where the primary and secondary geographic areas are spatially misaligned. However, the study by Bakar et al. [15] considered area-level data.

The aim of this research was to develop a hierarchical Bayesian spatial fusion small area model for binary data with spatial misalignment. The proposed model is a non-trivial extension of Bakar et al. [8] for the context of binary outcome variable under spatial misalignment with three contributions. First, the model uses unit-level survey data and area-level auxiliary data to predict the posterior mean proportion spatially at the second geographic area level. The individual process attempts to make better use of all the information gathered at the individual level. The area-level process, including the spatial random effect, which is estimated from the primary geographic areal units, is reused for prediction at the second area level. Second, the linking model is changed to $\text{logit}(p_i) = \log[p_i/(1-p_i)] \overset{iid}{\sim} N(\mu, \sigma^2)$, logit-normal model, in the proposed model. The logit-normal model facilitates easy computation of the parameters in the model. Last, the mean process was considered to overcome the multicollinearity between the true predictors and the spatial random effect. Unlike Bakar et al. [8], considering the mean process accounts for the hierarchical nature of the modeling.

The following is an outline of the remainder of this paper. Section 2 presents our modeling technique as well as the underlying distribution theory, backed up by relevant literature. Section 3 demonstrates the details of inference making under a number of sub-sections. Section 3's first two sub-sections are all about the details based on the primary geographic area, while the last two sub-sections are about the secondary geographic area. Section 4 highlights a simulation study illustrating how the proposed model can be put into practice. Concluding remarks are given in Section 5. Finally, appendices A–C present proofs and derivations of joint posterior, full conditional, and predictive distributions.

2. Model Development

This section introduces a spatial hierarchical Bayesian statistical small area model for the binary response variable. Let $Y_i(A_r)$ be a binary response variable for the ith individual in the rth area, A_r for $i = 1, \ldots, N_r, r = 1, \ldots, n$, and $\sum_{r=1}^{n} N_r = N$ where n denotes the number of small areas and N_r is the number of individuals in r^{th} area.

Consider $y_i(A_r)$ taking values one (with probabilities p_i) and zero (with probabilities $1 - p_i$), where $Y_i(A_r)$ follow Bernoulli distribution, that is, $y_i(A_r) \sim \text{ber}[p_i(A_r)]$. Hence, $y_i(A_r)$ is a Bernoulli response at area A_r and a logit model with probability $p_i(A_r)$ defined as follows:

$$\text{logit}[p_i(A_r)] = \mu(A_r) + \varepsilon_i(A_r). \tag{1}$$

This may also be written in a generalized linear mixed model form, where $g(\cdot)$ is considered as a logit link function, as follows:

$$g\{E(y_i(A_r))\} = \mu(A_r) + \varepsilon_i(A_r). \tag{2}$$

Assuming equal unit-level error variance [16], $\varepsilon_i(A_r) \sim N(0, \sigma_\varepsilon^2)$, and taking $X(A_r)$ as the area-level matrix of auxiliary covariates, then the first term of the right hand side of equation (2) becomes

$$\begin{aligned} \mu(A_r) &= \beta X(A_r) + \zeta(A_r) \\ &= \sum_{l=1}^{L} \beta_l x_l(A_r) + \zeta(A_r), \end{aligned} \tag{3}$$

where $x_l(A_r)$ is the lth auxiliary covariate at the area level and β_l is its respective regression coefficient with $l = 1, 2, \ldots, L$, where L refers to the number of auxiliary covariates in the model. We assume equal error variance, $\zeta(A_r) \sim N(0, \sigma_\zeta^2)$ [16], to make the posterior distributions proper.

The proposed hierarchical Bayesian spatial model for the binary data contains an underlying process whose predictive distribution on the probability scale is less uncertain than the observed proportions. The statistical dependencies caused by geographic proximity are modeled explicitly, and Bayesian conclusions account for them in a cohesive and natural way.

Therefore, embedding spatially correlated process,

$$\text{logit}\{p_i(A_r)\} = \mu(A_r) + \omega(A_r) + \varepsilon_i(A_r), \quad (4)$$

where the mean process, $\mu(A_r)$, is given in equation (3) and X is assumed not have a column for intercept, for the intercept to be implicitly present in spatial component [17]. This will not affect the model as the collinearity between the spatial random effects and the intercept is not of interest. Hence, we let the spatial random component remain unconstrained.

Furthermore, the term $\omega(A_r)$, the spatial component, is a spatially correlated process. Defining and understanding the spatial process $\omega(A_r) = (\omega(A_1), \ldots, \omega(A_n))'$ is typically essential for generating the full conditional distribution of ω.

Intrinsic conditional autoregressive (ICAR) and conditional autoregressive (CAR) models are popular methods used to model areal spatial random effects [18, 19], commonly considered in the Besag–York–Mollie (BYM) model. However, applying the CAR/ICAR technique to sparse area-level data is sometimes problematic since areas without observations are frequently neighbored by areas with observations. It is quite difficult to get spatial predictions in an overlapping area using the CAR modeling method. Another approach to model spatial random effect component is to use spatial basis functions. Moran's I basis function for a discrete Gaussian spatial field may be utilised to alleviate the sparseness and geographically overlapped predictions [20] by evaluating a latent spatial surface process formed by the basis.

Moran's I basis function [21] is one of the prominent techniques for modeling areal spatial process. In contrast to the CAR models, the process is updated using a multivariate distribution, which adds the big-n challenge stated in spatial literature [22]. As a result, for dimension reduction, m eigenvectors corresponding to the first m positive eigenvalues of a Moran's I operator matrix were used where $m < n$ [20]. Moran's I basis function, a discretely indexed areal spatial process model in general, which implements the same weights in geographical areas with large and small geographical border sizes, is frequently not appropriate [23]. However, due to the position and nature of the large area, it is feasible that the large area may contribute to a relatively minor spatial correlation with other surrounding areas. The other spatial basis function, the multiresolution basis function, similarly, could overcome the big-n problem.

The multiresolution spatial basis function is utilised for point-referenced data [24, 25]. In multiresolution spatial basis functions, m knot points are defined across the spatial domain [25] by utilising reduced rank approximation [26, 27]. Recently, a number of authors used spatial basis functions that included both the multiresolution basis function and Moran's I basis function [8, 15]. This integrates data from a neighborhood matrix of areas and distances between their centroids as well as allows non-stationarity [8].

This study used a spatial basis function for defining the spatial component, which included the multiresolution and Moran's I spatial basis functions for the spatial process. This study also developed a model-based strategy for handling the spatial predictive issue by employing a spatial basis function that tackles the sparse data problem via a latent spatial field and also accommodates the change of support problem. The spatial basis used in this study is a mix of Moran's I basis [21], which is used for area-level modeling, and the multiresolution spatial basis defined via the bi-square function, which is used for point-referenced spatial data [24, 28]. Such a basis function was also developed by Bakar [23] for a Gaussian model. However, this study extends further to the spatial generalized linear models.

Therefore, the spatial component of the model in equation (4), which can be expressed in terms of reduced spatial component (η) and the spatial basis function $\Phi_{(A)}$, is written as

$$\omega(A) = \Phi_{(A)}\eta + \nu, \quad (5)$$

where η is a vector of m dimension and $\Phi_{(A)}$ is a spatial basis that comprises Moran's I basis function and multiresolution basis function such that

$$\Phi_{(A)} = M_{(A)} \times R_{(A)}. \quad (6)$$

This is a mix of the multiresolution spatial basis $R_{(A)}$ and eigenvectors $M_{(A)}$ of Moran's I operator matrix $M(A)$. The term $\nu \sim N(\mathbf{0}, \sigma_\nu^2 \mathbf{I})$ in equation (5) is an identically and independently distributed error process used to capture the remaining random component so as to take the uncertainties arising from dimension reduction, $\Phi_{(A)}\eta$.

To derive spatial basis functions, there is a need to first derive both Moran's I basis function and multiresolution basis functions. For deriving Moran's I basis functions, Moran's I operator matrix, $M(A)$, from Moran's I statistic is derived. Moran's I statistic is a measure of spatial association, which equals a weighted sum of squares where the weights are called $M(A)$ [21]. For this, reparameterization is considered a gateway.

The reparameterization is used to demonstrate that the random effects can increase the variance of the posterior distribution of β due to the confounding between the spatial random effects and the fixed effect predictors (X) [17]. Reich et al. [17] defined the reparameterization as smoothing orthogonal to the fixed effects. Furthermore, Griffith et al. [29] augmented a model with selected eigenvectors of $(I - 11'/n)A(I - 11'/n)$ to reveal the structure of missing spatial covariates, where A is neighborhood weight matrix, 1 is the n-vector of 1s, and I is the $n \times n$ identity matrix. Observing that missing covariates are not of primary interest, rather in smoothing orthogonal to X. Hence, replace $I - 11'/n$ with $P^\perp$, where $P^\perp$ is the orthogonal projection onto $C(X)'s$ orthogonal complement, that is, $P^\perp = I - P$ with $P = X(X'X)^{-1}X'$. The resulting operator may be found in the numerator of Moran's I statistic, which is a popular non-parametric measure of spatial dependence [30].

Recently, the approach of dimension reduction by Reich et al. [17] was adapted by Bakar et al. [8, 15]. Bakar et al. [8]

used mean process for the case of modeling several categorical variables simultaneously unlike design matrix, X. In this study, the mean process generated from the predictors (μ) was used to construct the basis so as to alleviate the well-known problem of collinearity [17].

Considering the mean process, μ, generated from the predictors and neighborhood weight matrix, $W_{(A)}$, Moran's I operator matrix, is explicitly defined as

$$M(A) = \left(\mathbf{I} - \mu(\mu'\mu)^{-1}\mu'\right)W_{(A)}\left(\mathbf{I} - \mu(\mu'\mu)^{-1}\mu'\right), \qquad (7)$$

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

Instead of using the first m eigenvectors of $M(A)$, which correspond to the positive eigenvalues [21], n eigenvectors to define $M_{(A)}$ were employed [8, 17] for compatibility of matrix algebra. This may be accomplished by performing a spectral decomposition of Moran's I operator matrix $M(A)$ [20].

Set m knot points to define the multiresolution spatial basis $R_{(A)}$. The selection of the position and number of knots is a problem that is connected to $R_{(A)}$. For the position of knot points, a regular grid or its modification is generally a common choice [31, 32]. Then, $R_{(A)}$ of order $n \times m$ is produced using the distances between the knot points and the centroid of the areas, while a bi-square function is used to define the multiresolution spatial basis function, $R_{(A)}$. Therefore, for $j = 1, \ldots, r_i$ basis functions of the ith resolution, the bi-square basis function is defined in $\mathbb{R}^2$ as

$$S_j^{(i)}(A_r) = \left[1 - \left(\frac{\left\|A_r - c_j^{(i)}\right\|}{\phi_i}\right)^2\right]^2 I\left(\left\|A_r - c_j^{(i)}\right\| < \phi_i\right); A_r \in \mathbb{R}^2, \qquad (8)$$

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

The choice of $\{\phi_i\}$ determines the multiple resolutions, which are used to capture different dependence scales [24, 25, 27]. Several basis functions with centers beyond the research region are incorporated to account for boundary effects [33].

When knot-point sites are equidistant across the region, the value of ϕ_i is a multiple of the shortest distance between knot-point sites. Cressie et al. [24] utilised a constant range ϕ_i for all knot points with a particular resolution. To accommodate knot locations that may not be equidistant, the ranges of the knot points were enabled to vary. Hence, ϕ_i was defined as a multiple of the shortest distance between knot points and all the other knot points. In practice, ϕ_i is defined as the shortest Euclidean distance between basis function centers of the same resolution multiplied by 1.5 [24]. As a result, $\Phi_{(A)} = M_{(A)} \times R_{(A)}$ is a spatial basis that reflects the identification of the neighborhood phenomenon of the geographically specified areal borders as well as their distances.

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

$$\Sigma = \phi \times k(\cdot) = \phi \times \Phi'_{(A)}Q\Phi_{(A)}, \qquad (9)$$

where $\phi > 0$, Q is an $n \times n$ matrix with diagonal entries q_{ii} and non-diagonal entries. The diagonal entry is the number of region i's neighbors, and non-diagonal entries $q_{ij} = -1$ if areas j and i are neighbors, and 0 otherwise [17]. This allows one to incorporate neighborhood information into the covariance matrix, Σ. It can be computed using the neighborhood weight matrix, $W_{(A)}$, as $Q = \text{diag}(W_{(A)}1) - W_{(A)}$, and 1 is a vector of 1s.

Since Q is singular [34], a spectral decomposition of Σ was used to derive the full conditional distribution of the spatial process η.

Thus, we can write

$$\Sigma = \Psi(\phi \times \Lambda)\Psi s', \qquad (10)$$

where Λ is a diagonal matrix and Ψ is an orthogonal matrix of order $m \times m$ [20]. Following Hughes et al. [21], in this study, eigenvectors of Σ are assumed to be known with eigenvalues known up to a multiplicative constant.

Graphical representation of hierarchical Bayesian spatial small area models can be helpful for the readers in understanding structure and building complex models. Hence, the overall scheme of the model is given in Figure 1.

3. Bayesian Inference

3.1. Posterior Distribution.
This section describes the hierarchical derivation of the posterior distribution. The hierarchical specification for the desired joint model is in terms of data models, process models, and parameter models [35]. Such specification is likened to a simple fact from specifications of the Bayes theorem [35] (see Appendix A).

In Bayesian statistics, the posterior distribution has to be an appropriate distribution. It is vital to ensure that the sufficient conditions, in terms of the parameters' priors, are satisfied so that the resultant posterior is suitable. As a result, we applied proper priors to all unknown parameters to ensure that the posterior distribution is correct [36]. In choosing a prior belonging to a particular distributional family, some choices may be more computationally convenient than others. In particular, it may be possible to choose a prior distribution which is conjugate to the likelihood, in which case the prior and posterior distribution will be in the same family and have the same distributional form. Accordingly, a conjugate normal prior distribution for the elements of β and inverse gamma prior distribution for the variance parameters and smoothing parameter are used.

Direct assessment of the posterior distribution necessitates high-dimensional numerical integration, which is computationally infeasible. Therefore, the Markov Chain Monte Carlo (MCMC) algorithm was used to generate

samples from the posterior distributions [37]. The full conditional distributions are required to execute MCMC sampling under the model.

3.2. Full Conditional Distributions.

The full conditional distributions are derived by abstracting only those elements including the parameter of interest from the posterior distribution, in this case from equation (A.5), and treating other components as constants [38]. Accordingly, we obtained the full conditional distributions for the variance parameters (σ_ε^2, σ_ζ^2, and σ_ν^2), smoothing parameter (ϕ), and the coefficients of the fixed effects (β) of the model as follows:

$$
\begin{aligned}
\sigma_\varepsilon^2 | \cdot &\sim IG\left(\frac{N}{2} + a_\varepsilon, b_\varepsilon + \frac{1}{2} \sum_{r=1}^{n} \sum_{i=1}^{N_r} \left[\varphi_{ir} - \mu_r - \omega_r \right]^2 \right), \\
\sigma_\zeta^2 | \cdot &\sim IG\left(\frac{n}{2} + a_\zeta, b_\zeta + \frac{1}{2} \left[\mu(A) - \beta X(A) \right]\prime \left[\mu(A) - \beta X(A) \right] \right), \\
\sigma_\nu^2 | \cdot &\sim IG\left(\frac{n}{2} + a_\nu, b_\nu + \frac{1}{2} \left[\omega - \Phi_{(A)} \eta \right]\prime \left[\omega - \Phi_{(A)} \eta \right] \right), \\
\phi | \cdot &\sim IG\left(\frac{m}{2} + a_\phi, b_\phi + \frac{1}{2} \left[\eta\prime \Psi \Lambda^{-1} \Psi\prime \eta \right] \right).
\end{aligned}
\tag{11}
$$

Similarly, the conditional distribution of β can be denoted as $N(\triangle \chi, \triangle)$, i.e.,

$$
\beta | \cdot \sim N(\triangle \chi, \triangle), \tag{12}
$$

where

$$
\begin{aligned}
\triangle^{-1} &= \frac{1}{\sigma_\zeta^2} X' X + \frac{1}{\sigma_\beta^2} I_p, \\
\chi &= \frac{1}{\sigma_\zeta^2} X' \mu + \frac{1}{\sigma_\beta^2} \mu_\beta.
\end{aligned}
\tag{13}
$$

The proof for σ_ε^2, σ_ζ^2, and σ_ν^2 is straightforward. However, the corresponding proof for the derivation of conditional distribution of ϕ and β necessitates some algebra manipulation (see proofs in Appendix B). Therefore, since all the full conditional distributions have a closed form solution, samples are generated from the posterior distributions using the Gibbs sampling method [39].

3.3. Predictive Distributions.

Using estimates of the parameters under the primary geographic classification ($A_r, j = 1, \ldots, n$), we need to predict at area level with overlapping geographic boundaries. We assume that the areal relationships from A_r are largely maintained for the prediction area as they cover the same population and the same whole areas. This led to reusing all the parameters with few adjustment of spatial correlation according to the new geographical classification.

Let A_k^* with $k = 1, \ldots, n^*$ be the second geographical area where the ultimate prediction is required. Here we have a new neighborhood weight matrix ($W_{(A)}^*$) based on prediction areas. Let also X^* be the $n^* \times L$ known matrix of area-level auxiliary covariates based on second geographical boundaries. Hence, the mean process component, the spatial random component, and the individual-level error component corresponding to equation (4) for the prediction area will be computed separately and joined finally. Specifically, change of support is used to derive the probability distribution of the mean process component $\mu(A_k^*)$ at prediction areas (second geographical boundaries).

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

$$
\mu(A_k^*) = \frac{1}{|A_k^*|} \int_{A_r \subseteq A_k^*} \mu(A_r) \mathrm{d}A_r, \tag{14}
$$

where $|A_k^*|$ denotes the area of the areal unit A_k^*. However, the integral in equation (14) is difficult to compute, and there are approximation techniques available to resolve this problem [18]. However, the approximations are best suited for situations when $A_r \subseteq A_k^*$. These approximation techniques are derived using the concept of the posterior predictive distribution in the Bayesian inference approach (see Appendix C).

Thus, using the auxiliary covariates X^*, the posterior predictive distribution of $\mu(A_k^*)$ is written as

$$
\begin{aligned}
\pi\left(\mu(A_k^*) | \mu\right) &= \int \pi\left(\mu(A_k^*) | X^*, \beta, \sigma_\zeta^2, \mu\right) \times \pi\left(\beta, \sigma_\zeta^2 | X^*, \mu\right) \mathrm{d}\theta \\
&= E_{\left(\beta, \sigma_\zeta^2 | X^*, \mu\right)} \left[\pi\left(\mu(A_k^*) | X^*, \beta, \sigma_\zeta^2, \mu\right) \right].
\end{aligned}
\tag{15}
$$

On the other hand, for the spatial random effect, a new closest neighbor weight matrix $W_{(A^*)}$ based on $(A^*)_k's$ neighborhood in Moran's I-matrix and a new orthogonal multiresolution spatial basis $R_{(A^*)}$ based on the distances between knot points and $(A^*)_k s$ centroids are used. Thus, $\omega(A_k^*)$ is obtained by incorporating $\Phi_{(A^*)}$ in equation (5), where $\Phi^* = M_{(A^*)}^* \times R^*$ is the basis function containing the W^* neighborhood weight matrix generated from at A_k^*. $M_{(A^*)}$ and $R_{(A^*)}$, respectively, are the multiresolution basis and the eigenvectors of Moran's I operator matrix based on area A_k^*, i.e., SA2s.

Similarly, the term $\varepsilon_i(A_k^*)$ is derived from the predictive conditional Gaussian distribution with variance $\widehat{\sigma_\varepsilon^2}$ and mean 0. The predictive distribution of $y(A_k^*)$ is therefore produced using the composition of draws and MCMC samples from the posterior distributions and posterior predictive distribution. Hence, the predictive equation for t^{th} MCMC sample may be stated in terms of area A_k^* as

$$
\text{logit}\left(p_k^t\right) = \mu\left(A_k^*\right)^{(t)} + \omega\left(A_k^*\right)^{(t)} + \varepsilon_i\left(A_k^*\right)^{(t)}. \tag{16}
$$

3.4. Estimation and Evaluation Metrics.

Measures of precision play crucial role in small area estimation. For D as the number of MCMC samples after removing the burn-in period followed by thinning, by the ergodic theorem for Markov
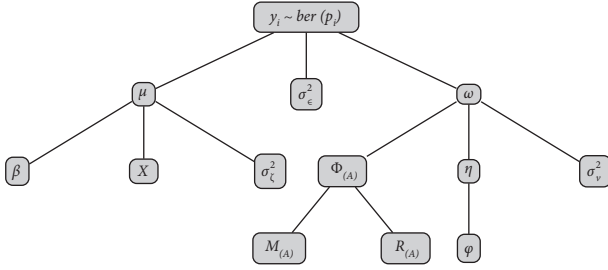
FIGURE 1: Directed acyclic graph (DAG) showing model schemes.

chains [40], $\widehat{p}_k$ converges to $E(p_k|y)$ and $\widehat{V}(p_k|y)$ to $V(p_k|y)$ as $D \longrightarrow \infty$. Hence, the estimate of p_i and its corresponding posterior variance for the kth area are obtained directly from the predictive distribution of $y(A_k^*)$. Accordingly,

$$\widehat{p}_k \approx \frac{1}{R} \sum_{t=1}^{R} p_k^{(t)} = p_k^{(\cdot)},$$

$$V(p_k \mid \widehat{p}) \approx \frac{1}{R-1} \sum_{t=1}^{R} \left(p_k^{(t)} - p_k^{(\cdot)} \right)^2. \tag{17}$$

We also considered mean absolute error (MAE) and root mean square error (RMSE) to evaluate the model performance. These metrics were defined using the posterior mean of the proportion, $\widehat{p}_{k\,d}$, of a small area k and iteration d from the true population proportion, p_k, in two ways: for an individual small area, and an average across all individual small areas. So, the most accurate estimate will lead to the smallest RMSE and/or MAE value going to zero.

The mean square error (MSE) was computed as the arithmetic mean of squared deviations of $\widehat{p}_{k\,d}$ from p_k, and hence the RMSE was the square root of MSE.

$$\mathrm{MSE}_k = \frac{1}{D} \sum_{d=1}^{D} \left(\widehat{p}_{kd} - p_k \right)^2,$$

$$\mathrm{MSE} = \frac{1}{n^*} \left(\sum_{k=1}^{n^*} \mathrm{MSE}_k \right). \tag{18}$$

Similarly, MAE is given as

$$\mathrm{MAE}_k = \frac{1}{D} \sum_{d=1}^{D} |p_{kd} - p_k|,$$

$$\mathrm{MAE} = \frac{1}{n^*} \left(\sum_{k=1}^{n^*} \mathrm{MAE}_k \right). \tag{19}$$

4. Illustrative Example

A simulation study was conducted as an illustrative example to make the ideas described in the proposed model clear and understand how the model may be put into practice. We started it by describing the simulation setup. The sensitivity analysis of the predictive posterior distribution for the

changes in the values of hyperparameters was done. Furthermore, precision measures of the estimates from the proposed model were demonstrated along with the direct estimates and population values.

4.1. Simulation Setup. A regular grids observation locations on a unit square, $[0, 1] \times [0, 1]$ with $20^2 = 400$ cells and $10^2 = 100$ cells, were considered as primary geographic areas with survey data and prediction area (secondary geographic area overlapping with primary geographic area), respectively (Figure 2(a)). A binary outcome variable and two auxiliary covariates with 3000 population observations were generated. Ten percent of the population was taken as sample observations.

Values for the parameters of the priors (i.e., hyperparameters) are assumed to reflect a fairly vague knowledge of the prior distributions. In case of inverse gamma distribution, smaller hyperparameters reflect vague knowledge. However, as noted in [41], inverse gamma priors with extremely small hyperparameters lack a proper limiting posterior distribution. As a result, we considered proper and weakly informative inverse gamma priors with reasonable values of the shape hyperparameter $a = 2$ and the rate hyperparameter $b = 1$, which is frequently recommended in literature [42, 43]. On the other hand, hyperparameters for the normal priors were considered under sensitivity analysis.

We generated an MCMC sample of 50,000 simulated datasets from the full conditionals of the hierarchical model. Bi-square basis function centers with two resolutions: resolution one containing $r_1 = 4 \times 4 = 16$ basis functions and resolution two containing $r_2 = 3 \times 3 = 9$ basis functions, were considered (Figure 2(b)). Therefore, in the simulation, there were $r = r_1 + r_2 = 25$ spatial basis functions in total.

4.2. Sensitivity Analysis. Sensitivity to the hyperparameter specification is a crucial part in a hierarchical Bayesian model. Assessing "complete factorial" sensitivity to our prior specifications is exceedingly difficult because of the model's complexity and size. However, it is suggested that sensitivity studies of one parameter at a time be performed by re-running the Gibbs sampler with various values for each parameter [44]. Hence, we investigated the sensitivity analysis for the normal prior. The sensitivities were mainly assessed by visual inspection of diagnostic plots (Figure 3) as well as using mean absolute error (MAE) and root mean square error (RMSE) (Table 1).

The model with the lowest RMSE and MAE is thought to perform better. RMSE and MAE were computed using different hyper-prior values from the table above, ranging from informative to vague (non-informative). At $\mu_\beta = 0$ and $\sigma_\beta^2 = 1$, the RMSE and MAE values are 0.43 and 0.40, respectively. Moving in the opposite direction, the next hyper-prior values are $\mu_\beta = 0$ and $\sigma_\beta^2 = 10$, with 0.02 and 0.03 increments, respectively. The step that follows produces the largest change increments of 0.04 and 0.06 for RMSE and MAE, respectively. After that, the increment becomes nearly constant. The RMSE and MAE values revealed that our model is not very sensitive to the hyperparameter selection;
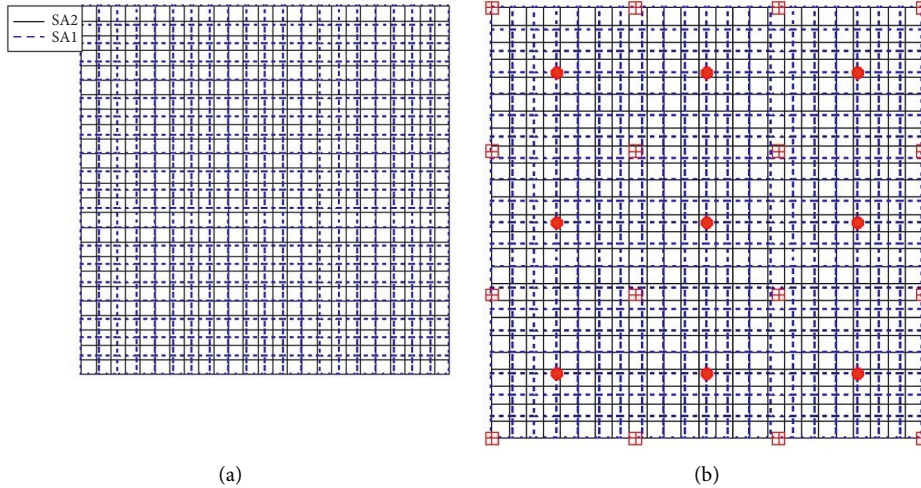
FIGURE 2: Regular grid cells and knot points. (a) Small area one (SA1) and small area two (SA2) overlaid together. (b) Knot points with two resolutions, red crossed square and red dotted circle overlaid on SA1 and SA2.
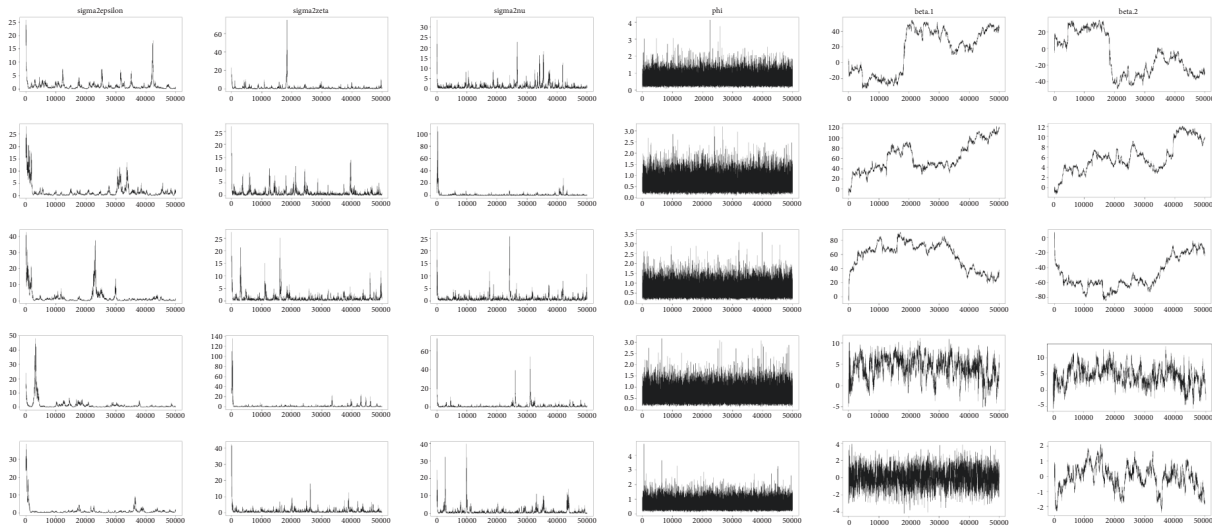

FIGURE 3: Trace plots for each parameter under five hyperparameters. Each row represents trace plots of parameters at a given hyperparameter, i.e., 1st row for $N(0, 10^{15})$, 2nd row for $N(0, 10^{10})$, 3rd row for $N(0, 10^5)$, 4th row for $N(0, 10)$, and 5th row for $N(0, 1)$. Each column represents trace plot of a parameter at each hyperparameter, i.e., 1st column for σ_ϵ^2, 2nd column for σ_ζ^2, 3rd column for σ_ν^2, 4th column for ϕ, 5th column for β_1, and 6th column for β_2.

TABLE 1: Sensitivity of the hyperparameters of the normal prior distribution.

Scenarios	Changes in hyperparameters	RMSE	MAE
Scenario 1	$\mu_\beta = 0$, $\sigma_\beta^2 = 10^{15}$	0.50	0.49
Scenario 2	$\mu_\beta = 0$, $\sigma_\beta^2 = 10^{10}$	0.49	0.49
Scenario 3	$\mu_\beta = 0$, $\sigma_\beta^2 = 10^5$	0.49	0.49
Scenario 4	$\mu_\beta = 0$, $\sigma_\beta^2 = 10$	0.45	0.43
Scenario 5	$\mu_\beta = 0$, $\sigma_\beta^2 = 1$	0.43	0.40

in other words, the observed changes in RMSE and MAE values were of a similar magnitude.

Visual inspection of the trace plots was also considered to complement analysis of RMSE and MAE (Figure 3). From this figure, we can observe that plots corresponding all parameters except β_1 and β_2 showed similar convergence.

Furthermore, the plots corresponding to a prior $N(0, 10)$ shows better convergence. This visual inspection was supported by Geweke's test of convergence for each parameter under each scenario (Table 2). Hence, we considered a normal prior with mean zero and variance ten, $N(0, 10)$, for our estimation and measures of precision.

4.3. Estimates and Precision Measures. Using samples from the MCMC algorithm and posterior predictive distribution along with the new information from SA2, estimates of probability and all precision measures were computed, including but not limited to 95% credible intervals (CIs). For comparison purpose, the direct estimate and the population values for each small areas were demonstrated.

TABLE 2: Geweke's test of convergence.

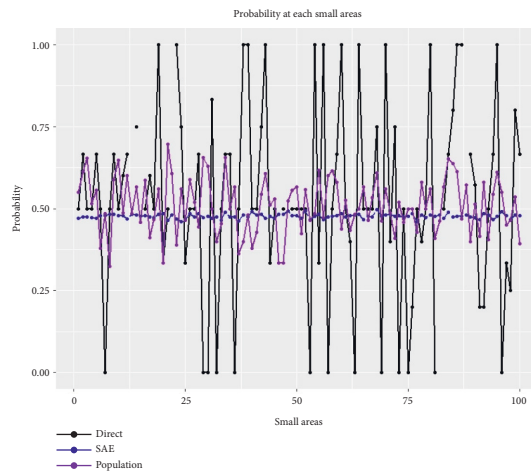| Scenarios	sigma2epsilon	sigma2zeta	sigma2nu	phi	beta.1	beta.2
Scenario 1	0.9286	1.4605	−0.2808	−1.1974	−6.4221	4.6221
Scenario 2	1.3933	1.2847	1.1283	0.6562	−2.4847	−1.8876
Scenario 3	2.2112	1.3935	0.6161	0.3807	−0.1418	−0.6526
Scenario 4	1.6848	0.7357	0.4744	1.1001	0.6030	−1.6672
Scenario 5	0.8739	0.9191	1.3838	−0.6712	0.3194	−1.2086

FIGURE 4: Probabilities (population, direct estimate, and small area estimate) of 100 small areas.
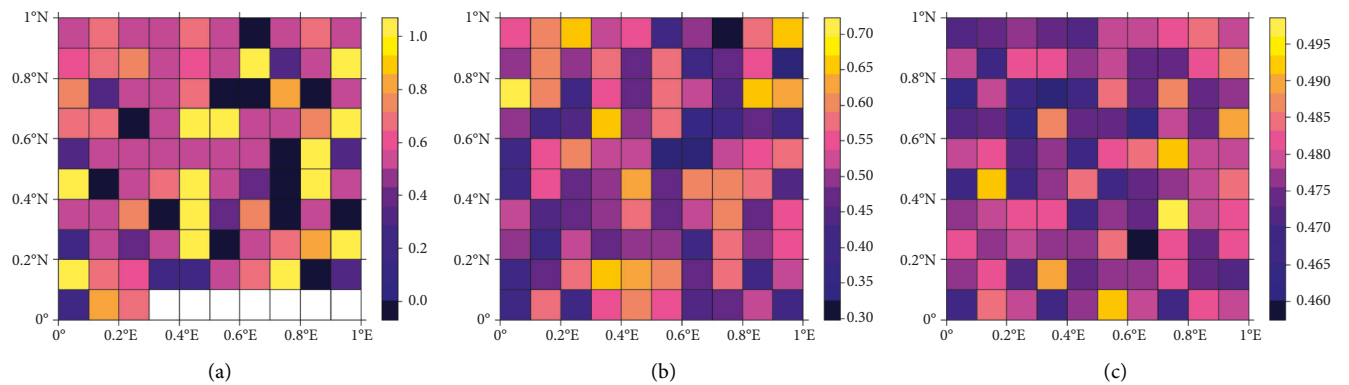

FIGURE 5: Population value and estimates over small areas. (a) Direct estimates. (b) The population/true value. (c) Estimate from the proposed model.

From Figures 4 and 5, we can see that our proposed model is better than direct estimate by considering several perspectives. Firstly, it enabled us to obtain estimates where we could not have from direct estimate, i.e., the line representing the direct estimates is broken for some areas. Secondly, if we look at the overall pattern of the lines, the deviation of the SAE line is less than the deviation of direct line from the population line. Hence, our proposed model yields estimates for areas where we could not have direct estimates and the estimates from our proposed model are precise as compared to the direct estimates.

From Figures 6 and 7, we can understand that the proposed model yields precise estimates as compared to direct estimates.

Different scenarios for the coefficients of auxiliary variables were considered to check the sensitivity of hyperparameters in the proposed model, and we found that weakly informative priors were better in the proposed model. However, it is recommended to conduct further studies assessing all combinations of hyperparameters in the proposed model. In this perspective, further research is needed.
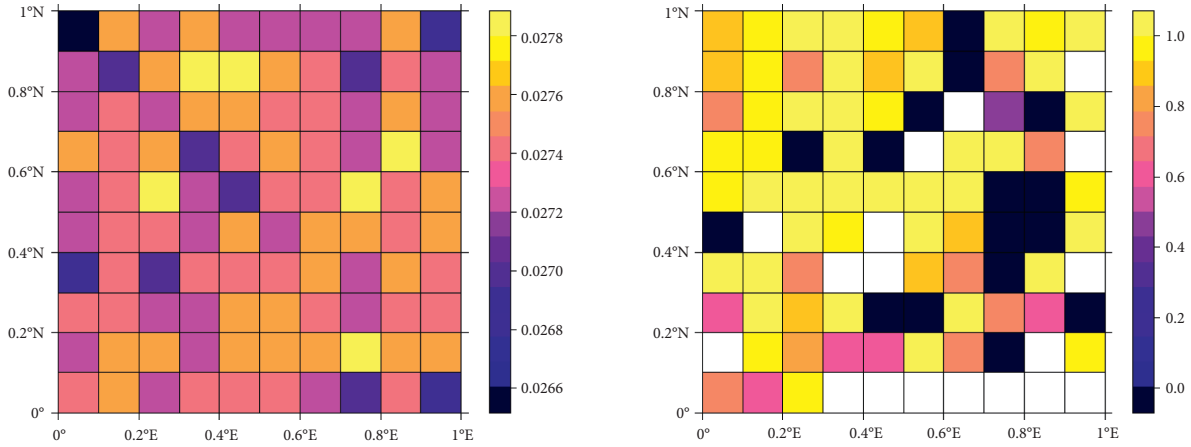
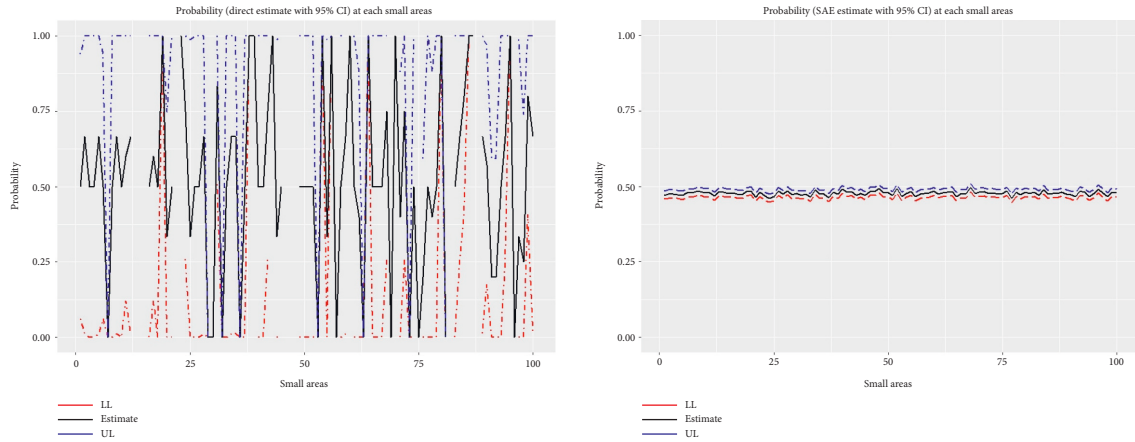FIGURE 6: The width of the 95 percent CI for small area estimates.

FIGURE 7: Probabilities (population, direct estimate, and small area estimate) of 100 small areas.

5. Concluding Remarks

In this study, we have developed a spatial hierarchical Bayesian small area model for a binary response variable under spatial misalignment as non-trivial extension of existing hierarchical Bayesian small area model for binary data under spatial misalignment. Fusion model (considering both area-level and unit-level data) was considered. The process models derived from the predictors were used to construct the basis as a means of alleviating the issue of collinearity between the fixed effects and the spatial random effects. All the full conditionals have closed form. Sensitivity analysis was done via simulation. The developed model can be applied in many broad areas including but not limited to health sciences, public health, agriculture, and economics.

Abbreviations

BYM: Besag–York–Mollie
CAR: Conditional autoregressive
CI: Credible interval
ICAR: Intrinsic conditional autoregressive
MAE: Mean absolute error
MCMC: Markov Chain Monte Carlo
RMSE: Root mean square error
SAE: Small area estimation.

Appendix

A. Posterior Distribution

For ease of expression, hereafter we denote $\varphi_{ir} = \text{logit} (p_i(A_r))$, $\mu_r = \mu(A_r)$, $\omega_r = \omega(A_r)$, $\mu = (\mu(A_1), \ldots, \mu(A_n))\prime$ and $\omega = (\omega(A_1), \ldots, \omega(A_n))\prime$. Hence, the conditional distributions under data models, process models, and parameter models used in posterior distribution are given as follows.

Data models:

$$y_i(A_r)|p_i(A_r) \sim \text{ber}(p_i(A_r)). \tag{A.1}$$

Process models:

$$
\begin{aligned}
\varphi_{ir}|\mu_r, \omega_r, \sigma_\varepsilon^2 &\sim N\left(\mu_r + \omega_r, \sigma_\varepsilon^2\right), \\
\mu|\beta, \sigma_\zeta^2 &\sim N\left(\beta X, \sigma_\zeta^2 I_n\right), \\
\omega|\eta, \sigma_\nu^2 &\sim N\left(\Phi\eta, \sigma_\nu^2 I_n\right), \\
\eta|\phi &\sim N(0, \Sigma).
\end{aligned}
\tag{A.2}
$$

Parameter models:

$$\beta \sim N\left(\mu_\beta, \sigma_\beta^2 I_P\right),$$
$$\sigma_\varepsilon^2 \sim IG\left(a_\varepsilon, b_\varepsilon\right),$$
$$\sigma_\zeta \sim IG\left(a_\zeta, b_\zeta\right), \qquad (A.3)$$
$$\sigma_\nu^2 \sim IG\left(a_\nu, b_\nu\right),$$
$$\phi \sim IG\left(a_\phi, b_\phi\right),$$

where $\sigma_\varepsilon^2 > 0; \ \sigma_\zeta^2 > 0; \ \sigma_\nu^2 > 0; \ \phi > 0; \ i = 1, 2, \ldots, N_r$, and $r = 1, 2, \ldots, n$.

Accordingly, the posterior distribution is written as

$$\pi\left(\mu, \omega, \eta, \beta, \sigma_\varepsilon^2, \sigma_\nu^2, \sigma_\zeta^2, \phi | y_i\right) \propto \prod_{r=1}^{n} \prod_{i=1}^{N_r} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left\{-\frac{1}{2} \frac{\left[\varphi_{ir} - \mu_r - \omega_r\right]^2}{\sigma_\varepsilon^2}\right\}$$

$$\times \frac{1}{(2\pi)^{n/2}\sigma_\zeta^n} \exp\left\{-\frac{1}{2} \frac{\left(\mu - X\beta\right)'\left[\mu - X\beta\right]}{\sigma_\zeta^2}\right\}$$

$$\times \frac{1}{(2\pi)^{n/2}\sigma_\nu^n} \exp\left\{-\frac{1}{2} \frac{\left[\omega - \Phi\eta\right]'\left[\omega - \Phi\eta\right]}{\sigma_\nu^2}\right\}$$

$$\times (2\pi)^{-(m/2)}|\Sigma|^{-(1/2)} \exp\left\{-\frac{1}{2}\eta'\Sigma^{-1}\eta\right\}$$

$$\times \frac{1}{(2\pi)^{p/2}\sigma_\beta^p} \exp\left\{-\frac{1}{2} \frac{\left[\beta - \mu_\beta\right]'\left[\beta - \mu_\beta\right]}{\sigma_\beta^2}\right\}$$

$$\times \frac{b_\varepsilon^{a_\varepsilon}}{\Gamma(a_\varepsilon)}\left(\sigma_\varepsilon^2\right)^{-a_\varepsilon-1} \exp\left(-\frac{b_\varepsilon}{\sigma_\varepsilon^2}\right) \times \frac{b_\zeta^{a_\zeta}}{\Gamma(a_\zeta)}\left(\sigma_\zeta^2\right)^{-a_\zeta-1} \exp\left(-\frac{b_\zeta}{\sigma_\zeta^2}\right)$$

$$\times \frac{b_\nu^{a_\nu}}{\Gamma(a_\nu)}\left(\sigma_\nu^2\right)^{-a_\nu-1} \exp\left(-\frac{b_\nu}{\sigma_\nu^2}\right) \times \frac{b_\phi^{a_\phi}}{\Gamma(a_\phi)}\left(\phi\right)^{-a_\phi-1} \exp\left(-\frac{b_\phi}{\phi}\right)$$

$$\times \prod_{r=1}^{n} \prod_{i=1}^{N_r} p_i(A_r)^{y_i(A_r)} \left(1 - p_i(A_r)\right)^{1 - y_i(A_r)},$$

$$(A.4)$$

which is proportional to a product of the kernels only after removing terms that do not involve the parameter of interest that can be written as

$$\pi\left(\mu, \omega, \eta, \beta, \sigma_\varepsilon^2, \sigma_\nu^2, \sigma_\zeta^2, \phi | y_i\right) \propto \prod_{r=1}^{n} \prod_{i=1}^{N_r} \frac{1}{\sigma_\varepsilon} \exp\left\{-\frac{1}{2} \frac{\left[\varphi_{ir} - \mu_r - \omega_r\right]^2}{\sigma_\varepsilon^2}\right\}$$

$$\times \frac{1}{\sigma_\zeta^n} \exp\left\{-\frac{1}{2} \frac{\left(\mu - X\beta\right)'\left[\mu - X\beta\right]}{\sigma_\zeta^2}\right\}$$

$$\times \frac{1}{\sigma_\nu^n} \exp\left\{-\frac{1}{2} \frac{\left[\omega - \Phi\eta\right]'\left[\omega - \Phi\eta\right]}{\sigma_\nu^2}\right\}$$

$$\times |\Sigma|^{-(1/2)} \exp\left\{-\frac{1}{2}\eta'\Sigma^{-1}\eta\right\} \times \frac{1}{\sigma_\beta^p} \exp\left\{-\frac{1}{2} \frac{\left[\beta - \mu_\beta\right]'\left[\beta - \mu_\beta\right]}{\sigma_\beta^2}\right\} \quad (A.5)$$

$$\times \left(\sigma_\varepsilon^2\right)^{-a_\varepsilon-1} \exp\left(-\frac{b_\varepsilon}{\sigma_\varepsilon^2}\right) \times \left(\sigma_\zeta^2\right)^{-a_\zeta-1} \exp\left(-\frac{b_\zeta}{\sigma_\zeta^2}\right)$$

$$\times \left(\sigma_\nu^2\right)^{-a_\nu-1} \exp\left(-\frac{b_\nu}{\sigma_\nu^2}\right) \times \left(\phi\right)^{-a_\phi-1} \exp\left(-\frac{b_\phi}{\phi}\right)$$

$$\times \prod_{r=1}^{n} \prod_{i=1}^{N_r} p_i(A_r)^{y_i(A_r)} \left(1 - p_i(A_r)\right)^{1 - y_i(A_r)}.$$

B. Derivation of Full Conditionals

(1) The full conditionals for σ_ϵ^2.

  The full conditionals for σ_ϵ^2 can be proved by taking all terms with σ_ϵ^2 from equation (A.5) as

$$\pi\left(\sigma_\varepsilon^2 | \cdot\right) \propto \prod_{r=1}^{n} \prod_{i=1}^{N_r} \frac{1}{\sigma_\varepsilon} \exp\left\{-\frac{1}{2} \frac{\left[\varphi_{ir} - \mu_r - \omega_r\right]^2}{\sigma_\varepsilon^2}\right\} \times \left(\sigma_\varepsilon^2\right)^{-a_\varepsilon-1} \exp\left(-\frac{b_\varepsilon}{\sigma_\varepsilon^2}\right)$$

$$\propto \left[\left(\sigma_\varepsilon^2\right)^{-a_\varepsilon-1} \prod_{r=1}^{n} \prod_{i=1}^{N_r} \frac{1}{\sigma_\varepsilon}\right]\left[\exp\left(-\frac{b_\varepsilon}{\sigma_\varepsilon^2}\right) \prod_{r=1}^{n} \prod_{i=1}^{N_r} \exp\left\{-\frac{1}{2} \frac{\left[\varphi_{ir} - \mu_r - \omega_r\right]^2}{\sigma_\varepsilon^2}\right\}\right]$$

$$\propto \left[\left(\sigma_\varepsilon^2\right)^{-a_\varepsilon-1}\left(\frac{1}{\sigma_\varepsilon}\right)^{-N}\right]\left[\exp\left(-\frac{b_\varepsilon}{\sigma_\varepsilon^2}\right)\exp\sum_{r=1}^{n} \sum_{i=1}^{N_r}\left\{-\frac{1}{2} \frac{\left[\varphi_{ir} - \mu_r - \omega_r\right]^2}{\sigma_\varepsilon^2}\right\}\right] \qquad (B.1)$$

$$\propto \left[\left(\sigma_\varepsilon^2\right)^{-a_\varepsilon-1}\left(\sigma_\varepsilon^2\right)^{-(N/2)}\right]\left[\exp\left(-\frac{b_\varepsilon}{\sigma_\varepsilon^2}\right)\exp\left\{\left(-\frac{1}{\sigma_\varepsilon^2}\right)\frac{1}{2}\sum_{r=1}^{n} \sum_{i=1}^{N_r}\left[\varphi_{ir} - \mu_r - \omega_r\right]^2\right\}\right]$$

$$\propto \left[\left(\sigma_\varepsilon^2\right)^{-\left(\left(a_\varepsilon+N\right)/2\right)-1}\right]\left[\exp\left\{\left(-\frac{1}{\sigma_\varepsilon^2}\right)\left(b_\varepsilon + \frac{1}{2}\sum_{r=1}^{n} \sum_{i=1}^{N_r}\left[\varphi_{ir} - \mu_r - \omega_r\right]^2\right)\right\}\right],$$

which is the kernel of inverse gamma distribution with shape and scale parameters

$$a_\varepsilon + \frac{N}{2}$$

$$\left(b_\varepsilon + \frac{1}{2} \sum_{r=1}^{n} \sum_{i=1}^{N_r} [\varphi_{ir} - \mu_r - \omega_r]^2 \right), \tag{B.2}$$

respectively.

$$\Rightarrow \pi(\sigma_\varepsilon^2|\cdot) \sim IG\left(\alpha_\varepsilon + \frac{N}{2}, b_\varepsilon + \frac{1}{2} \sum_{r=1}^{n} \sum_{i=1}^{N_r} [\varphi_{ir} - \mu_r - \omega_r]^2 \right). \tag{B.3}$$

(2) The full conditionals for σ_ζ^2.

The full conditionals for σ_ζ^2 can be proved by taking all terms with σ_ζ^2 from equation (A.5) as

$$\begin{aligned}
\pi(\sigma_\zeta^2|\cdot) &\propto \frac{1}{\sigma_\zeta^n} \exp\left\{ -\frac{1}{2} \frac{(\mu - X\beta)'[\mu - X\beta]}{\sigma_\zeta^2} \right\} \times (\sigma_\zeta^2)^{-a_\zeta - 1} \exp\left(-\frac{b_\zeta}{\sigma_\zeta^2} \right) \\
&\propto \frac{1}{\sigma_\zeta^n} (\sigma_\zeta^2)^{-a_\zeta - 1} \exp\left\{ -\frac{1}{2} \frac{(\mu - X\beta)'[\mu - X\beta]}{\sigma_\zeta^2} \right\} \exp\left(-\frac{b_\zeta}{\sigma_\zeta^2} \right) \\
&\propto (\sigma_\zeta^2)^{-(n/2) - a_\zeta - 1} \exp\left\{ -\frac{1/2[\mu - X\beta]'[\mu - X\beta] + b_\zeta}{\sigma_\zeta^2} \right\} \\
&\propto (\sigma_\zeta^2)^{-(n/2 + a_\zeta) - 1} \exp\left\{ -\frac{1/2[\mu - X\beta]'[\mu - X\beta] + b_\zeta}{\sigma_\zeta^2} \right\},
\end{aligned} \tag{B.4}$$

which has the form of inverse gamma distribution with a shape parameter of

$$\left(\frac{n}{2} + a_\zeta \right), \tag{B.5}$$

and scale parameter of

$$\frac{1}{2}[\mu - X\beta]'[\mu - X\beta] + b_\zeta, \tag{B.6}$$

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

(3) The full conditionals for σ_ν^2.

The full conditionals for σ_ν^2 can be proved by taking all terms with σ_ν^2 from equation (A.5) as

$$\begin{aligned}
\pi(\sigma_\nu^2|\cdot) &\propto \frac{1}{\sigma_\nu^n} \exp\left\{ -\frac{1}{2} \frac{[\omega - \Phi\eta]'[\omega - \Phi\eta]}{\sigma_\nu^2} \right\} \times (\sigma_\nu^2)^{-a_\nu - 1} \exp\left(-\frac{b_\nu}{\sigma_\nu^2} \right) \\
&\propto (\sigma_\nu^2)^{-(n/2 + a_\nu) - 1} \exp\left\{ -\frac{1/2[\omega - \Phi\eta]'[\omega - \Phi\eta] + b_\nu}{\sigma_\nu^2} \right\}.
\end{aligned} \tag{B.7}$$

Hence, the full conditional distribution of σ_ν^2 has the form of inverse gamma distribution with scale parameter

$$\frac{1}{2}[\omega - \Phi\eta]'[\omega - \Phi\eta] + b_\nu, \tag{B.8}$$

and shape parameter

$$\left(\frac{n}{2} + a_\nu \right)$$

$$\Rightarrow \pi(\sigma_\nu^2|\cdot) \sim IG\left(\frac{n}{2} + a_\nu, \frac{1}{2}[\omega - \Phi\eta]'[\omega - \Phi\eta] + b_\nu \right). \tag{B.9}$$

(4) The full conditionals for ϕ.

The full conditionals for ϕ can be proved by taking all terms with ϕ from the full joint posterior distribution equation (A.5) as

$$\pi(\phi|\cdot) \propto |\Sigma|^{-(1/2)} \exp\left\{-\frac{1}{2}\eta'\Sigma^{-1}\eta\right\} \times \phi^{-a-1} \exp\left\{-\frac{b}{\phi}\right\}. \quad (B.10)$$

Here, applying properties of determinant and inverse of a matrix as well as the relationship $\Sigma = \Psi(\phi\Lambda)\Psi\prime$, we get

$$\begin{aligned} |\Sigma|^{-(1/2)} &= \phi^{-(m/2)}|\Psi\Lambda\Psi'|^{-(1/2)}, \\ \Sigma^{-1} &= \phi^{-1}\Psi\Lambda^{-1}\Psi', \end{aligned} \quad (B.11)$$

implying

$$\begin{aligned} \pi(\phi|\cdot) &\propto \phi^{-(m/2)} \exp\left\{-\frac{1}{2}\phi^{-1}\eta'\Psi\Lambda^{-1}\Psi'\eta\right\} \times \phi^{-a-1} \exp\left\{-\frac{b}{\phi}\right\} \\ &\propto \phi^{-(m/2)}\phi^{-a-1} \exp\left\{-\frac{\eta'\Psi\Lambda^{-1}\Psi'\eta}{2\phi} + \frac{b}{\phi}\right\} \\ &\propto \phi^{-((m/2)+a)-1} \exp\left\{-\frac{1}{\phi}\left[\frac{\eta'\Psi\Lambda^{-1}\Psi'\eta}{2} + b\right]\right\}, \end{aligned} \quad (B.12)$$

which is inverse gamma distribution with parameters

$$\begin{aligned} &\frac{m}{2} + a_\phi, \\ &\frac{1}{2}\eta'\Psi\Lambda^{-1}\Psi'\eta + b_\phi, \quad (B.13) \\ \Rightarrow &\pi(\phi|\cdot) \sim IG\left(\frac{m}{2} + a_\phi, \frac{1}{2}\eta'\Psi\Lambda^{-1}\Psi'\eta + b_\phi\right). \end{aligned}$$

(5) The full conditionals for β.

The corresponding proof for the conditional distribution of β is as follows. Taking all terms with the β from the full joint posterior distribution equation (A.5),

$$\begin{aligned} \pi(\beta|\cdot) &\propto \exp\left\{-\frac{1}{2}\frac{[\mu - X\beta]'[\mu - X\beta]}{\sigma_\zeta^2}\right\} \times \exp\left\{-\frac{1}{2}\frac{[\beta - \mu_\beta]'[\beta - \mu_\beta]}{\sigma_\beta^2}\right\} \\ &\propto \exp\left\{-\frac{1}{2}\frac{[\mu - X\beta]'[\mu - X\beta]}{\sigma_\zeta^2} - \frac{1}{2}\frac{[\beta - \mu_\beta]'[\beta - \mu_\beta]}{\sigma_\beta^2}\right\} \\ &\propto \exp\left\{-\frac{1}{2}\frac{\mu'\mu - 2\beta'X'\mu + \beta'X'X\beta}{\sigma_\zeta^2} - \frac{1}{2}\frac{\beta'\beta - 2\beta'\mu_\beta + \mu_\beta'\mu_\beta}{\sigma_\beta^2}\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{-2\beta'X'\mu + \beta'X'X\beta}{\sigma_\zeta^2} + \frac{\beta'\beta - 2\beta'\mu_\beta}{\sigma_\beta^2}\right]\right\} \propto \exp\left\{-\frac{1}{2}\left[\frac{\beta'X'X\beta}{\sigma_\zeta^2} + \frac{\beta'\beta}{\sigma_\beta^2} - \frac{2\beta'X'\mu}{\sigma_\zeta^2} - \frac{2\beta'\mu_\beta}{\sigma_\beta^2}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\beta'\left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2\sigma_\beta^2}\right)\beta - 2\beta'\left(\frac{\sigma_\beta^2 X'\mu + \sigma_\zeta^2\mu_\beta}{\sigma_\zeta^2\sigma_\beta^2}\right)\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2\sigma_\beta^2}\right)\left[\beta\prime\beta - 2\beta\prime\left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2\sigma_\beta^2}\right)^{-1}\left(\frac{\sigma_\beta^2 X'\mu + \sigma_\zeta^2\mu_\beta}{\sigma_\zeta^2\sigma_\beta^2}\right)\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2\sigma_\beta^2}\right)\left[\beta - \left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2\sigma_\beta^2}\right)^{-1}\left(\frac{\sigma_\beta^2 X'\mu + \sigma_\zeta^2\mu_\beta}{\sigma_\zeta^2\sigma_\beta^2}\right)\right]'\left[\beta - \left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2\sigma_\beta^2}\right)^{-1}\left(\frac{\sigma_\beta^2 X'\mu + \sigma_\zeta^2\mu_\beta}{\sigma_\zeta^2\sigma_\beta^2}\right)\right]\right\}, \end{aligned} \quad (B.14)$$

which is multivariate normal distribution with covariance

$$\left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2 \sigma_\beta^2} \right)^{-1}, \tag{B.15}$$

and mean vector

$$\left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2 \sigma_\beta^2} \right)^{-1} \left(\frac{\sigma_\beta^2 X'\mu + \sigma_\beta^2 \mu_\beta}{\sigma_\zeta^2 \sigma_\beta^2} \right),$$

$$\Rightarrow \pi(\beta|\cdot) \sim N\left(\left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2 \sigma_\beta^2} \right)^{-1} \left(\frac{\sigma_\beta^2 X'\mu + \sigma_\beta^2 \mu_\beta}{\sigma_\zeta^2 \sigma_\beta^2} \right), \left(\frac{\sigma_\beta^2 X'X + \sigma_\zeta^2 I_p}{\sigma_\zeta^2 \sigma_\beta^2} \right)^{-1} \right). \tag{B.16}$$

C. Predictive Distribution

The posterior predictive distribution is the distribution of potential unobserved values conditional on the observed values. The posterior predictive distribution for unobserved values (S') conditional on the observed values (S) is defined as follows:

$$p(S'|S) = \int_\theta p(S'|\theta)p(\theta|S). \tag{C.1}$$

This can be derived from the Bayes rules:

$$p(S', \theta|S) = p(S'|\theta, S)p(\theta|S). \tag{C.2}$$

We have

$$p(S', \theta|S) = \frac{p(S', \theta, S)}{p(S)} = \frac{p(S'\theta, S)p(\theta, S)}{p(S)} = p(S'|\theta, S)p(\theta|S). \tag{C.3}$$

Now integrating out the nuisance variable θ on both the most left and right sides in equation (C.3) yields

$$p(S'|S) = \int_\theta p(S'|\theta, S)p(\theta|S)\mathrm{d}\theta. \tag{C.4}$$

The right hand side formula in equation (C.4) appears to have a Markov-type assumption, $p(S'|\theta, S) = p(S'|\theta)$. Alternatively, by the law of total expectation and Fubini's theorem [45], applied to any bounded measurable function f defined on the relevant sample space Ω, we observe that

$$\int_\Omega f(S')p(S'|S)\mathrm{d}S' = E[f(S')|S = S]$$

$$= E[E[f(S')|\Theta, S]|S = S],$$

$$\text{Since, } E(X) = E(E(X|Y)) = \int_\theta \left(\int_\Omega f(S')p(S'|\theta, S)\mathrm{d}S' \right) p(\theta|S)\mathrm{d}\theta$$

$$= \int_\Omega f(S')\left(\int_\theta p(S'|\theta, S)p(\theta|S)\mathrm{d}\theta \right)\mathrm{d}S', \tag{C.5}$$

$$\Rightarrow \int_\Omega f(S') \underbrace{p(S'|S)}\, \mathrm{d}S' = \int_\Omega f(S')\left(\int_\theta \underbrace{p(S'|\theta, S)p(\theta|S)\mathrm{d}\theta} \right)\mathrm{d}S'.$$

Obviously, the under-braced components of the left hand side and the right hand side in equation (C.5) are equal,

and hence it is equivalent to equation (C.4). That is, for all bounded measurable functions, we conclude that

$$p(S'|S) = \int_{\theta} p(S'|\theta, S)p(\theta|S)\mathrm{d}\theta. \qquad \text{(C.6)}$$

Now, replacing S' by μ_k^*, S by μ_k and θ by β, σ_ζ^2 in equation (C.6), we find the predictive posterior distribution of μ_k^*.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

The views expressed in this publication are those of the authors. The funder had no role in manuscript writing, editing, approval, or decision to publish.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

KFM and SMM were responsible for conceptualization and software. KFM, SMM, and AKW were responsible for methodology, validation, investigation, original draft preparation, review and editing, and visualization. SMM and AKW were responsible for supervision. KFM was responsible for formal analysis, resources, data curation, project administration, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

References

[1] J. N. Rao and I. Molina, *Small Area Estimation*, John Wiley & Sons, New York, NY, USA, 2015.

[2] G. Datta and M. Ghosh, "Small area shrinkage estimation," *Statistical Science*, vol. 27, no. 1, pp. 95–114, 2012.

[3] M. Ghosh, "Hierarchical and empirical Bayes multivariate estimation," in *Current Issues in Statistical Inference (Lecture Notes-Monograph Series),* Institute of Mathematical Statistics, New York, NY, USA, 1992.

[4] J. N. Rao and I. Molina, "Introduction," in *Small Area Estimation,* John Wiley & Sons, New York, NY, USA, 2015.

[5] C. E. Särndal, B. Swensson, and J. Wretman, *Model Assisted Survey Sampling*, Springer Science & Business Media, Berlin, Germany, 2003.

[6] R. Lehtonen and A. Veijanen, "Design-based methods of estimation for domains and small areas," in *Handbook of Statistics*, vol. 29, pp. 219–249, Elsevier, Amsterdam, Netherlands, 2009.

[7] J. Breidenbach, S. Magnussen, J. Rahlf, and R. Astrup, "Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data," *Remote Sensing of Environment*, vol. 212, pp. 199–211, 2018.

[8] K. Shuvo Bakar, N. Biddle, P. Kokic, and H. Jin, "A bayesian spatial categorical model for prediction to overlapping geographical areas in sample surveys," *Journal of the Royal Statistical Society: Series A*, vol. 183, no. 2, pp. 535–563, 2020.

[9] S. K. Sahu, A. E. Gelfand, and D. M. Holland, "Fusing point and areal level spacea time data with application to wet deposition," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 59, no. 1, pp. 77–103, 2010.

[10] D. Pfeffermann, "New important developments in small area estimation," *Statistical Science*, vol. 28, no. 1, pp. 40–68, 2013.

[11] M. Ghosh, K. Natarajan, T. W. F. Stroud, and B. P. Carlin, "Generalized linear models for small-area estimation," *Journal of the American Statistical Association*, vol. 93, no. 441, pp. 273–282, 1998.

[12] A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, *Handbook of Spatial Statistics*, CRC Press, Boca Raton, FL, USA, 2010.

[13] M. D. Ugarte, "Banerjee, S. Carlin, B. P., and gelfand, A. E. Hierarchical modeling and analysis for spatial data. Second edition. CRC press/chapman & Hall. Monographs on statistics and applied probability 135, boca raton, Florida, 2015. 562 pp. $ 99.95 . ISBN-13: 978-1," *Biometrics*, vol. 71, no. 1, pp. 274–277, 2015.

[14] M. Trevisani and A. Gelfand, "Spatial misalignment models for small area estimation: a simulation study," in *Advances in Theoretical and Applied Statistics,* Springer, Berlin, Germany, 2013.

[15] K. S. Bakar and H. Jin, "Areal prediction of survey data using Bayesian spatial generalised linear models," *Communications in Statistics—Simulation and Computation*, vol. 49, no. 11, pp. 2963–2978, 2018.

[16] Y. You and J. Rao, "Hierarchical Bayes estimation of small area means using multi-level models," *Survey Methodology*, vol. 26, no. 2, pp. 173–181, 2000.

[17] B. J. Reich, J. S. Hodges, and V. Zadnik, "Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models," *Biometrics*, vol. 62, no. 4, pp. 1197–1206, 2006.

[18] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, CRC Press, Boca Raton, FL, USA, 2014.

[19] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, CRC Press, Boca Raton, FL, USA, 2005.

[20] J. R. Bradley, C. K. Wikle, and S. H. Holan, "Bayesian spatial change of support for count-valued survey data with application to the American community survey," *Journal of the American Statistical Association*, vol. 111, no. 514, pp. 472–487, 2016.

[21] J. Hughes and M. Haran, "Dimension reduction and alleviation of confounding for spatial generalized linear mixed models," *Journal of the Royal Statistical Society: Series B*, vol. 75, no. 1, pp. 139–159, 2013.

[22] S. K. Sahu and K. Bakar, "A comparison of bayesian models for daily ozone concentration levels," *Statistical Methodology*, vol. 9, no. 1-2, pp. 144–157, 2012.

[23] K. S. Bakar, "Bayesian gaussian models for interpolating large-dimensional data at misaligned areal units," 2017, https://arxiv.org/abs/1711.03666.

[24] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B*, vol. 70, no. 1, pp. 209–226, 2008.

[25] D. Nychka, S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain, "A multiresolution Gaussian process model for the analysis of large spatial datasets," *Journal of Computational & Graphical Statistics*, vol. 24, no. 2, pp. 579–599, 2015.

[26] N. Cressie, T. Shi, and E. L. Kang, "Fixed rank filtering for spatio-temporal data," *Journal of Computational & Graphical Statistics*, vol. 19, no. 3, pp. 724–745, 2010.

[27] M. Katzfuss, "A multi-resolution approximation for massive spatial datasets," *Journal of the American Statistical Association*, vol. 112, no. 517, pp. 201–214, 2017.

[28] E. L. Kang and N. Cressie, "Bayesian inference for the spatial random effects model," *Journal of the American Statistical Association*, vol. 106, no. 495, pp. 972–983, 2011.

[29] D. A. Griffith, *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*, Springer, Berlin, Germany, 2003.

[30] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, p. 17, 1950.

[31] P. Diggle and S. Lophaven, "Bayesian geostatistical design," *Scandinavian Journal of Statistics*, vol. 33, no. 1, pp. 53–64, 2006.

[32] S. K. Sahu, K. Shuvo Bakar, and N. Awang, "Bayesian forecasting using spatiotemporal models with applications to ozone concentration levels in the Eastern United States," *Geometry Driven Statistics*, vol. 121, pp. 260–281, 2015.

[33] N. Cressie and E. L. Kang, "High-resolution digital soil mapping: kriging for very large datasets," in *Proximal Soil Sensing*, pp. 49–63, Springer, Berlin, Germany, 2010.

[34] P. J. Diggle and P. J. Ribeiro, *Model Based Geostatistics*, Springer-Verlag, New York, NY, USA, 2007.

[35] L. M. Berliner, "Hierarchical Bayesian time series models," in *Maximum Entropy and Bayesian Methods*, pp. 15–22, Springer, Berlin, Germany, 1996.

[36] J. P. Hobert and G. Casella, "The effect of improper priors on Gibbs sampling in hierarchical linear mixed models," *Journal of the American Statistical Association*, vol. 91, no. 436, pp. 1461–1473, 1996.

[37] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[38] W. R. Gilks, "Full conditional distributions," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. Spiegelhalter, Eds., Chapman & Hall, New York, NY, USA, 1996.

[39] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

[40] S. Orey, "An ergodic theorem for Markov chains," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 1, no. 2, pp. 174–176, 1962.

[41] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper)," *Bayesian Analysis*, vol. 1, no. 3, pp. 515–534, 2006.

[42] A. Gelman, A. Jakulin, M. G. Pittau, and Y. S. Su, "A weakly informative default prior distribution for logistic and other regression models," *Annals of Applied Statistics*, vol. 2, no. 4, pp. 1360–1383, 2008.

[43] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, CRC Press, Boca Raton, FL, USA, 2013.

[44] C. K. Wikle, R. F. Milliff, D. Nychka, and L. M. Berliner, "Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 382–397, 2001.

[45] J. Mikusiński, "The Fubini theorem," in *The Bochner Integral* Springer, Berlin, Germany, 1978.