Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial Bayesian Modeling of Child Malnutrition with Measurement Error in Demographic Health Survey Data

Romuald Daniel BOY-NGBOGBELE

romuald.boy-ngbogbele@students.jkuat.ac.ke

Pan African University

Bertin Dehigbe

Institute of Mathematics and Physical Sciences

Aboubakar Alfa Samuel

Institute of Mathematics and Physical Sciences

Research Article

Keywords: Spatial Bayesian modeling, Measurement error, Childhood malnutrition, Demographic and Health Surveys (DHS), Hierarchical logistic regression, Spatial epidemiology

Posted Date: April 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9369506/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spatial Bayesian Modeling of Child Malnutrition with Measurement Error in Demographic Health Survey Data

Romuald Daniel Boy-ngbogbele [*1], Bertin Dehigbe [†2], and Aboubakar Alfa Samuel [‡2]

[1]Department of Mathematics, Pan African University Institute for Basic Science Technology and Innovation, Nairobi, Kenya, romuald.boy-ngbogbele@students.jkuat.ac.ke
[2]Department of Statistics and Probability, Institute of Mathematics and Physical Sciences, University of Abomey-Calavi , Porto-Novo, Benin

Abstract

Background: Childhood malnutrition remains a major public health challenge in low- and middle-income countries, particularly in Sub-Saharan Africa. Conventional statistical models often ignore spatial dependence and measurement error, potentially leading to biased estimates and reduced predictive performance. This study proposes a spatial Bayesian logistic regression model that incorporates both measurement error and spatial dependence for analyzing pooled cross-sectional data. A simulation study, designed to mimic Demographic and Health Survey (DHS) data, was conducted using multiple clusters and survey periods. Covariates were generated with realistic measurement error, and spatial correlation was introduced through a distance-based covariance structure. The performance of the spatial model was compared with a non-spatial Bayesian model using accuracy, area under the curve (AUC), precision, recall, F1-score, and estimated prevalence. The spatial Bayesian model consistently outperformed the non-spatial model across all evaluation metrics. Improvements were observed in accuracy (approximately 3–5%), AUC (0.05–0.07 increase), precision, recall, and F1-score. Additionally, the spatial model produced lower and more stable prevalence estimates, indicating improved calibration and reduced bias. These results demonstrate the advantages of incorporating spatial structure in modeling childhood malnutrition. Incorporating spatial dependence and measurement error within a Bayesian framework significantly enhances model performance and provides more reliable estimates of childhood malnutrition. The proposed approach offers a valuable tool for improving health data analysis and supports more effective targeting of public health interventions.

Keywords: Spatial Bayesian modeling; Measurement error; Childhood malnutrition; Demographic and Health Surveys (DHS); Hierarchical logistic regression; Spatial epidemiology.

1 Introduction

Childhood malnutrition remains one of the most serious global public health challenges, particularly in low- and middle-income countries (LMICs). Despite significant progress in nutrition programs and

[*]Corresponding author: R. D. B. romuald.boy-ngbogbele@students.jkuat.ac.ke
[†]bertin.dehigbe@imsp-uac.org
[‡]aboubakar.alfa@imsp-uac.org

international development initiatives, millions of children under five continue to suffer from stunting, wasting, and underweight conditions. Recent global estimates indicate that more than 148 million children are affected by stunting worldwide, with the highest prevalence occurring in Sub-Saharan Africa and South Asia (UNICEF, 2023; WHO, 2024). Malnutrition during early childhood has severe long-term consequences, including impaired cognitive development, reduced educational attainment, and increased risk of chronic diseases later in life (Black et al., 2022; Victora et al., 2023). Consequently, reducing childhood malnutrition remains a major priority in global health and sustainable development agendas.

Large-scale household surveys such as the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) have become critical sources of data for investigating child health outcomes in LMICs. These surveys provide nationally representative information on child nutrition indicators, household socioeconomic characteristics, and maternal health conditions. The availability of repeated cross-sectional survey waves allows researchers to examine temporal trends and identify key determinants of child malnutrition across countries and regions (Pullum et al., 2022; Croft et al., 2023). In addition, the integration of geospatial information within DHS datasets has enabled researchers to conduct spatial analyses of health outcomes and explore geographic disparities in nutrition and health indicators (Burgert et al., 2022; Macharia et al., 2023).

Numerous empirical studies have investigated the socioeconomic and demographic determinants of childhood malnutrition using conventional statistical models. Logistic regression and multilevel mixed-effects models are commonly applied to identify risk factors such as household wealth, maternal education, sanitation conditions, and dietary practices (Akombi et al., 2022; Amare et al., 2023). Recent studies have also incorporated contextual variables such as community infrastructure, regional food security, and access to healthcare services in order to better understand the multidimensional drivers of malnutrition (Hagos et al., 2023; Tesema et al., 2024). Although these models have provided valuable insights into the determinants of malnutrition, they often assume that explanatory variables are measured without error.

Measurement error in explanatory variables is a common problem in health survey data and can lead to biased parameter estimates and misleading statistical inference. Many variables collected in demographic and health surveys rely on self-reported information, including breastfeeding frequency, dietary diversity, and household socioeconomic conditions. Such variables may be affected by recall bias, reporting errors, or interviewer effects (Carroll et al., 2022; Keogh et al., 2023). Ignoring measurement error in regression models may attenuate estimated associations between risk factors and health outcomes, resulting in underestimation of the true effects of important determinants (Gustafson, 2022; Yi et al., 2024). Therefore, statistical models that explicitly account for measurement error are essential for producing reliable estimates in health research.

Bayesian statistical methods provide a flexible framework for addressing measurement error through hierarchical modeling and latent variable approaches. In Bayesian models, the true values of covariates can be treated as latent variables and estimated jointly with other model parameters, allowing uncertainty from measurement error to be propagated throughout the entire inferential process (Carroll et al., 2022; Yi et al., 2024). Recent methodological developments have demonstrated the advantages of Bayesian frameworks for improving the accuracy of statistical inference in health survey data, particularly when dealing with pooled cross-sectional datasets that combine multiple survey waves (Boy-ngbogbele et al., 2026; Liu et al., 2023). For example, Boy-ngbogbele et al. (2026) proposed a Bayesian framework designed to enhance health data accuracy in pooled cross-sectional analyses by explicitly modeling measurement error and survey heterogeneity. Such approaches highlight the importance of integrating measurement error correction within hierarchical Bayesian models when analyzing

large-scale health survey data.

Another important characteristic of childhood malnutrition is its strong spatial heterogeneity. Health outcomes often vary across geographic regions due to differences in environmental conditions, food availability, health infrastructure, and socioeconomic development. Spatial analyses have shown that malnutrition tends to cluster geographically, creating regional hotspots where children face disproportionately high risks of undernutrition (Macharia et al., 2023; Amare et al., 2023). Understanding these spatial patterns is essential for designing targeted public health interventions and allocating resources efficiently to the most vulnerable populations (Hagos et al., 2023; Tesema et al., 2024).

Spatial statistical models provide powerful tools for analyzing geographic variation in health outcomes. Bayesian spatial models, including Conditional Autoregressive (CAR) models and spatial hierarchical regression models, are widely used in disease mapping and spatial epidemiology (Lawson, 2022; Moraga, 2023). These models account for spatial autocorrelation between neighboring regions and allow researchers to identify geographic clusters of disease or health outcomes. Recent applications of spatial modeling in nutrition research have revealed substantial geographic disparities in malnutrition prevalence across African and Asian countries (Moraga, 2023; Macharia et al., 2023).

Although spatial modeling has become increasingly common in epidemiological research, relatively few studies have integrated spatial dependence with measurement error correction in a unified statistical framework. Most existing analyses either focus on spatial variation in health outcomes or address measurement error in regression models separately (Gustafson, 2022; Liu et al., 2023). However, failing to simultaneously account for spatial dependence and measurement error may lead to biased estimates and reduced predictive performance, particularly in analyses involving complex survey data with hierarchical and geographic structures.

Recent advances in Bayesian hierarchical modeling provide new opportunities to address these challenges by integrating spatial effects and measurement error correction within a single modeling framework. Bayesian spatial hierarchical models allow the simultaneous modeling of latent variables, spatial dependence, and random effects at multiple levels (Lawson, 2022; Moraga, 2023). Furthermore, improvements in computational methods such as Markov Chain Monte Carlo (MCMC) algorithms have made it feasible to estimate complex hierarchical models using large-scale survey datasets (Nandram et al., 2024; Yi et al., 2024). These methodological developments open new avenues for improving statistical analyses of malnutrition data.

Motivated by these developments, this study proposes a spatial Bayesian logistic regression model with measurement error for analyzing childhood malnutrition using Demographic and Health Survey data. The proposed model extends conventional hierarchical logistic regression by incorporating spatial random effects to capture geographic dependence while simultaneously correcting for measurement error in key explanatory variables. By integrating spatial modeling and measurement error correction within a unified Bayesian framework, the proposed approach aims to provide more accurate parameter estimates, improve predictive performance, and offer deeper insights into the geographic determinants of childhood malnutrition.

Research Gaps and Motivation

Despite the extensive literature on childhood malnutrition and its determinants, important methodological limitations remain in the statistical analysis of health survey data. Many empirical studies rely on conventional regression models that assume explanatory variables are measured without error and that observations are independent across geographic regions. However, demographic and health surveys often contain measurement inaccuracies due to recall bias, reporting errors, and survey design

limitations. Ignoring such measurement error can lead to biased parameter estimates and misleading conclusions regarding the determinants of malnutrition.

Recent research has emphasized the usefulness of Bayesian hierarchical models for addressing measurement error in health survey data. For example, Boy-ngbogbele et al. (2026) proposed a Bayesian framework for improving health data accuracy in pooled cross-sectional analyses by explicitly modeling measurement error and survey heterogeneity. While this approach represents an important advancement for correcting inaccuracies in survey-based health data, it does not account for spatial dependence in malnutrition outcomes. At the same time, spatial epidemiological studies have demonstrated that childhood malnutrition often exhibits strong geographic clustering due to regional differences in socioeconomic conditions, access to health services, and environmental factors.

These limitations highlight an important methodological gap in the existing literature, as few studies simultaneously address both measurement error and spatial dependence in the analysis of childhood malnutrition. Motivated by this gap, the present study proposes a spatial Bayesian logistic regression model with measurement error for analyzing childhood malnutrition using DHS data. By integrating spatial random effects with measurement error correction within a unified hierarchical Bayesian framework, the proposed model aims to improve parameter estimation, capture geographic heterogeneity, and provide more reliable evidence for designing targeted nutrition policies.

2 Materials and Methods

2.1 Study Design

This study employs a quantitative observational design using pooled cross-sectional data from the Demographic and Health Surveys (DHS). DHS surveys are nationally representative household surveys that collect information on population health, nutrition, and socioeconomic conditions in low- and middle-income countries. The analysis focuses on children under five years of age and combines multiple DHS survey waves into a pooled dataset to increase sample size and improve statistical power while allowing the examination of temporal variations in childhood malnutrition.

The DHS surveys use a two-stage stratified cluster sampling design. In the first stage, enumeration areas are selected using probability proportional to size, and in the second stage, households are systematically selected within each cluster. Information on child health, anthropometric measurements, maternal characteristics, and household socioeconomic conditions is collected through standardized questionnaires. The primary outcome variable in this study is a binary indicator of childhood malnutrition derived from anthropometric indices such as height-for-age, weight-for-height, and weight-for-age.

To investigate the determinants of childhood malnutrition, the study incorporates a range of child-level, maternal-level, and household-level covariates, including child age, breastfeeding practices, parental education, household wealth index, and place of residence. Geographic coordinates associated with DHS clusters are also utilized to account for spatial dependence in malnutrition outcomes. A spatial Bayesian logistic regression model with measurement error is applied to capture cluster-level variability, temporal effects across survey years, measurement error in selected covariates, and spatial heterogeneity across regions.

2.2 Study Setting

This study was conducted in Cameroon, a lower-middle-income country located in Central Africa. Cameroon is characterized by substantial geographic, socioeconomic, and environmental diversity, which

influences health outcomes across the country. The nation is divided into ten administrative regions: Adamawa, Centre, East, Far North, Littoral, North, Northwest, West, South, and Southwest. These regions differ significantly in terms of population density, access to healthcare services, infrastructure development, and food security conditions.

Cameroon has a mixed economy with large disparities between urban and rural areas. Rural populations often face limited access to healthcare facilities, clean water, sanitation infrastructure, and adequate nutrition. These inequalities contribute to variations in child health outcomes, including malnutrition. The Far North, North, and Adamawa regions in particular have historically recorded higher levels of childhood malnutrition due to poverty, climate vulnerability, and food insecurity, while relatively lower prevalence rates are often observed in the southern and urbanized regions of the country.

The analysis utilizes data from multiple waves of the Demographic and Health Surveys conducted in Cameroon, which provide nationally representative information on child health, nutrition, and household socioeconomic characteristics. The DHS surveys also provide geographic coordinates for survey clusters, allowing spatial analysis of health outcomes across regions. The diverse geographic and socioeconomic context of Cameroon makes it an appropriate setting for investigating spatial patterns of childhood malnutrition and evaluating the role of measurement error and spatial heterogeneity in statistical modeling of health survey data.

2.2.1 Formulation of the Spatial Bayesian Logistic Regression Model with Measurement Error

This study extends the Bayesian logistic regression model with measurement error by incorporating spatial dependence across geographic regions. The modeling framework is developed progressively, beginning with the baseline logistic regression model, followed by the inclusion of measurement error, hierarchical random effects, and finally spatial dependence.

2.2.2 Baseline Logistic Regression Model

Let Y_{ijt} denote a binary indicator of malnutrition for a child i in a cluster j during the survey year t. The baseline logistic regression model is defined as

$$Y_{ijt} \sim \text{Bernoulli}(p_{ijt}) \tag{1}$$

where $p_{ijt} = P(Y_{ijt} = 1)$ represents the probability that child i is malnourished.
The logistic regression specification is given by

$$\text{logit}(p_{ijt}) = \log\left(\frac{p_{ijt}}{1 - p_{ijt}}\right) = \beta_0 + \beta_1 X_{1ijt} + \beta_2 X_{2ijt} + \cdots + \beta_k X_{kijt} \tag{2}$$

where

- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_k$ are regression coefficients.
- $X_{1ijt}, X_{2ijt}, \ldots, X_{kijt}$ represent explanatory variables describing child, household, and socioeconomic characteristics (Boy-ngbogbele et al., 2026).

2.2.3 Measurement Error Model

In demographic and health surveys, several covariates may be measured with error due to recall bias, reporting inaccuracies, or data collection limitations. Let X^*_{kijt} denote the observed covariate measured with error, while X_{kijt} represents the true latent value (Boy-ngbogbele et al., 2026).

The classical additive measurement error model is defined as

$$X^*_{kijt} = X_{kijt} + v_{kijt} \tag{3}$$

where

$$v_{kijt} \sim \mathcal{N}(0, \sigma^2_v) \tag{4}$$

represents the measurement error term.

The logistic regression model incorporating measurement error becomes

$$\text{logit}(p_{ijt}) = \beta_0 + \beta_1 X^*_{1ijt} + \beta_2 X^*_{2ijt} + \cdots + \beta_k X^*_{kijt} \tag{5}$$

2.2.4 Hierarchical Random Effects

Children within the same survey cluster may share similar environmental or socioeconomic conditions. In addition, malnutrition prevalence may vary across survey years. To account for this hierarchical structure, cluster-level and time-level random effects are introduced (Boy-ngbogbele et al., 2026).

The hierarchical logistic regression model becomes

$$\text{logit}(p_{ijt}) = \beta_0 + \sum_{k=1}^{K} \beta_k X^*_{kijt} + u_j + w_t \tag{6}$$

where

- u_j represents cluster-level random effects capturing unobserved heterogeneity across clusters.
- w_t represents time-level random effects capturing temporal variability.

These random effects are assumed to follow

$$u_j \sim \mathcal{N}(0, \sigma^2_u) \tag{7}$$

$$w_t \sim \mathcal{N}(0, \sigma^2_w) \tag{8}$$

2.2.5 Incorporating Spatial Dependence

Malnutrition outcomes may exhibit geographic clustering due to regional differences in infrastructure, health services, climate conditions, and socioeconomic environments. To capture spatial dependence, a spatially structured random effect s_j is introduced.

The spatial Bayesian logistic regression model becomes

$$\text{logit}(p_{ijt}) = \beta_0 + \sum_{k=1}^{K} \beta_k X^*_{kijt} + s_j + u_j + w_t \tag{9}$$

where s_j represents the spatial effect associated with region or cluster j.

2.2.6 Conditional Autoregressive (CAR) Spatial Model

The spatial random effects are modeled using a Conditional Autoregressive (CAR) prior, which assumes that neighboring regions have similar risk levels.

Let $N(j)$ denote the set of neighboring regions of region j, and let n_j represent the number of neighbors. The CAR prior is defined as

$$s_j | s_{-j}, \tau_s \sim \mathcal{N}\left(\frac{1}{n_j}\sum_{l \in N(j)} s_l, \frac{1}{n_j \tau_s}\right) \tag{10}$$

where

- s_{-j} denotes the spatial effects of all regions except j,
- τ_s is the spatial precision parameter controlling the strength of spatial dependence.

This formulation encourages neighboring regions to have similar spatial effects, thereby capturing geographic clustering of malnutrition.

2.2.7 Full Hierarchical Spatial Bayesian Model

The complete hierarchical spatial Bayesian model is summarized as follows:

$$Y_{ijt} \sim \text{Bernoulli}(p_{ijt}) \tag{11}$$

$$\text{logit}(p_{ijt}) = \beta_0 + \sum_{k=1}^{K} \beta_k X^*_{kijt} + s_j + u_j + w_t \tag{12}$$

$$X^*_{kijt} = X_{kijt} + v_{kijt}, \quad v_{kijt} \sim \mathcal{N}(0, \sigma_v^2) \tag{13}$$

$$u_j \sim \mathcal{N}(0, \sigma_u^2) \tag{14}$$

$$w_t \sim \mathcal{N}(0, \sigma_w^2) \tag{15}$$

$$s_j \sim CAR(\tau_s) \tag{16}$$

2.2.8 Prior Distributions

To complete the Bayesian model, prior distributions are specified as follows:

$$\beta_k \sim \mathcal{N}(0, 1000) \tag{17}$$

$$\sigma_u^2, \sigma_w^2, \sigma_v^2 \sim \text{Inverse-Gamma}(a, b) \tag{18}$$

$$\tau_s \sim \text{Gamma}(a_s, b_s) \tag{19}$$

The model parameters are estimated using Markov Chain Monte Carlo (MCMC) simulation methods.

2.2.9 Posterior Distribution of the Spatial Bayesian Model

Bayesian inference is based on the posterior distribution obtained from Bayes' theorem. The posterior distribution combines the likelihood of the observed data with the prior distributions of the model parameters.

2.2.10 Likelihood Function

Let Y_{ijt} denote the binary malnutrition outcome for child i in cluster or region j at time t. Conditional on the probability p_{ijt}, the outcome follows a Bernoulli distribution:

$$Y_{ijt} \sim \text{Bernoulli}(p_{ijt}) \tag{20}$$

The likelihood function for the entire dataset is therefore

$$L(\mathbf{Y}|\boldsymbol{\beta}, \mathbf{X}^*, \mathbf{s}, \mathbf{u}, \mathbf{w}) = \prod_{i=1}^{n} \prod_{j=1}^{J} \prod_{t=1}^{T} p_{ijt}^{Y_{ijt}} (1 - p_{ijt})^{1-Y_{ijt}} \tag{21}$$

where

$$p_{ijt} = \frac{\exp(\eta_{ijt})}{1 + \exp(\eta_{ijt})} \tag{22}$$

and

$$\eta_{ijt} = \beta_0 + \sum_{k=1}^{K} \beta_k X^*_{kijt} + s_j + u_j + w_t \tag{23}$$

represents the linear predictor.

2.2.11 Measurement Error Model

The observed covariates X^*_{kijt} are assumed to be contaminated with measurement error. The relationship between the observed and true covariates is given by

$$X^*_{kijt} = X_{kijt} + v_{kijt} \tag{24}$$

where

$$v_{kijt} \sim \mathcal{N}(0, \sigma_v^2) \tag{25}$$

The likelihood associated with the measurement error model is therefore

$$p(X^*_{kijt}|X_{kijt}, \sigma_v^2) = \prod_{i,j,t,k} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(X^*_{kijt} - X_{kijt})^2}{2\sigma_v^2}\right) \tag{26}$$

2.2.12 Prior Distributions

Prior distributions are assigned to all unknown parameters in the hierarchical model.

2.2.13 Regression coefficients

$$\beta_k \sim \mathcal{N}(0, \sigma_\beta^2), \quad k = 0, 1, \ldots, K \tag{27}$$

Cluster random effects

$$u_j \sim \mathcal{N}(0, \sigma_u^2), \quad j = 1, \ldots, J \tag{28}$$

Time random effects

$$w_t \sim \mathcal{N}(0, \sigma_w^2), \quad t = 1, \ldots, T \tag{29}$$

Spatial effects

The spatial effects follow a Conditional Autoregressive (CAR) prior:

$$s_j | s_{-j}, \tau_s \sim \mathcal{N}\left(\frac{1}{n_j} \sum_{l \in N(j)} s_l, \frac{1}{n_j \tau_s}\right) \tag{30}$$

where $N(j)$ denotes the set of neighboring regions and n_j is the number of neighbors. The joint distribution of the spatial effects can be written as

$$p(\mathbf{s}|\tau_s) \propto \tau_s^{(J-1)/2} \exp\left(-\frac{\tau_s}{2} \sum_{j \sim l} (s_j - s_l)^2\right) \tag{31}$$

where $j \sim l$ indicates neighboring regions.

Variance parameters

Inverse-Gamma priors are assigned to the variance components:

$$\sigma_u^2 \sim \text{Inverse-Gamma}(a_u, b_u) \tag{32}$$

$$\sigma_w^2 \sim \text{Inverse-Gamma}(a_w, b_w) \tag{33}$$

$$\sigma_v^2 \sim \text{Inverse-Gamma}(a_v, b_v) \tag{34}$$

The spatial precision parameter follows

$$\tau_s \sim \text{Gamma}(a_s, b_s) \tag{35}$$

Joint Posterior Distribution

Let

$$\Theta = (\boldsymbol{\beta}, \mathbf{s}, \mathbf{u}, \mathbf{w}, \sigma_v^2, \sigma_u^2, \sigma_w^2, \tau_s, \mathbf{X})$$

denote the set of all unknown parameters and latent variables.

Using Bayes' theorem, the joint posterior distribution is proportional to the product of the likelihood and prior distributions:

$$p(\Theta|\mathbf{Y}, \mathbf{X}^*) \propto p(\mathbf{Y}|\boldsymbol{\beta}, \mathbf{X}^*, \mathbf{s}, \mathbf{u}, \mathbf{w})\; p(\mathbf{X}^*|\mathbf{X}, \sigma_v^2)\; p(\boldsymbol{\beta})\; p(\mathbf{s}|\tau_s)\; p(\mathbf{u}|\sigma_u^2)\; p(\mathbf{w}|\sigma_w^2)\; p(\sigma_v^2)\; p(\sigma_u^2)\; p(\sigma_w^2)\; p(\tau_s) \tag{36}$$

Substituting the likelihood and prior expressions yields

$$p(\Theta|\mathbf{Y}, \mathbf{X}^*) \propto \prod_{i,j,t} p_{ijt}^{Y_{ijt}} (1 - p_{ijt})^{1 - Y_{ijt}} \tag{37}$$

$$\times \prod_{i,j,t,k} \exp\left(-\frac{(X^*_{kijt} - X_{kijt})^2}{2\sigma_v^2} \right) \tag{38}$$

$$\times \prod_{k=0}^{K} \exp\left(-\frac{\beta_k^2}{2\sigma_\beta^2} \right) \tag{39}$$

$$\times \exp\left(-\frac{\tau_s}{2} \sum_{j \sim l} (s_j - s_l)^2 \right) \tag{40}$$

$$\times \prod_{j=1}^{J} \exp\left(-\frac{u_j^2}{2\sigma_u^2} \right) \tag{41}$$

$$\times \prod_{t=1}^{T} \exp\left(-\frac{w_t^2}{2\sigma_w^2} \right) \tag{42}$$

$$\times p(\sigma_v^2) p(\sigma_u^2) p(\sigma_w^2) p(\tau_s) \tag{43}$$

Posterior Inference

The posterior distribution does not have a closed-form solution due to the hierarchical structure and spatial dependence. Therefore, parameter estimation is performed using Markov Chain Monte Carlo (MCMC) methods such as Gibbs sampling or the Metropolis–Hastings algorithm.

2.3 Posterior Computation Using MCMC Methods

The joint posterior distribution derived in the previous section does not have a closed-form solution due to the nonlinear logistic link, the hierarchical structure, and the spatial dependence introduced by the Conditional Autoregressive (CAR) model. Consequently, posterior inference is performed using Markov Chain Monte Carlo (MCMC) simulation methods.

In this study, a hybrid MCMC algorithm combining Gibbs sampling and the Metropolis–Hastings (MH) algorithm is employed to generate samples from the posterior distribution.

2.3.1 Overview of the MCMC Algorithm

Let

$$\Theta = (\boldsymbol{\beta}, \mathbf{s}, \mathbf{u}, \mathbf{w}, \sigma_v^2, \sigma_u^2, \sigma_w^2, \tau_s, \mathbf{X})$$

denote the set of all unknown parameters and latent variables. The MCMC algorithm iteratively samples each parameter from its full conditional distribution while conditioning on the most recent values of the remaining parameters.

At iteration $m+1$, the parameter vector is updated sequentially as

$$\Theta^{(m+1)} = (\boldsymbol{\beta}^{(m+1)}, \mathbf{s}^{(m+1)}, \mathbf{u}^{(m+1)}, \mathbf{w}^{(m+1)}, \sigma_v^{2(m+1)}, \sigma_u^{2(m+1)}, \sigma_w^{2(m+1)}, \tau_s^{(m+1)}, \mathbf{X}^{(m+1)}).$$

2.3.2 Gibbs Sampling Steps

Parameters with conjugate prior distributions can be updated directly using Gibbs sampling.

2.3.3 Updating the cluster random effects

The cluster effects u_j follow a normal prior distribution and their conditional posterior distribution is proportional to

$$p(u_j|\cdot) \propto p(\mathbf{Y}|\beta, X^*, s_j, u_j, w_t)\; p(u_j|\sigma_u^2)$$

which results in a normal full conditional distribution:

$$u_j|\cdot \sim \mathcal{N}(\mu_{u_j}, \sigma_{u_j}^2).$$

2.3.4 Updating the time effects

Similarly, the temporal random effects are updated as

$$w_t|\cdot \sim \mathcal{N}(\mu_{w_t}, \sigma_{w_t}^2).$$

2.3.5 Updating variance parameters

Given the inverse-gamma priors, the variance parameters have conjugate full conditional distributions:

$$\sigma_u^2|\cdot \sim \text{Inverse-Gamma}\left(a_u + \frac{J}{2}, b_u + \frac{1}{2}\sum_{j=1}^{J} u_j^2\right) \tag{44}$$

$$\sigma_w^2|\cdot \sim \text{Inverse-Gamma}\left(a_w + \frac{T}{2}, b_w + \frac{1}{2}\sum_{t=1}^{T} w_t^2\right) \tag{45}$$

$$\sigma_v^2|\cdot \sim \text{Inverse-Gamma}\left(a_v + \frac{nK}{2}, b_v + \frac{1}{2}\sum_{i,j,t,k} (X_{kijt}^* - X_{kijt})^2\right) \tag{46}$$

2.3.6 Updating the spatial precision parameter

The spatial precision parameter τ_s also has a conjugate conditional posterior:

$$\tau_s|\cdot \sim \text{Gamma}\left(a_s + \frac{J-1}{2}, b_s + \frac{1}{2}\sum_{j\sim l}(s_j - s_l)^2\right). \tag{47}$$

2.4 Metropolis–Hastings Updates

Certain parameters do not have closed-form full conditional distributions due to the logistic likelihood. These parameters are updated using the Metropolis–Hastings algorithm.

2.4.1 Updating regression coefficients

The regression coefficients $\boldsymbol{\beta}$ are updated using a random-walk Metropolis step.

A proposal value $\boldsymbol{\beta}'$ is generated from

$$\boldsymbol{\beta}' \sim \mathcal{N}(\boldsymbol{\beta}^{(m)}, \Sigma_\beta)$$

where Σ_β is a proposal covariance matrix.

The acceptance probability is

$$\alpha = \min\left(1, \frac{p(\mathbf{Y}|\boldsymbol{\beta}', X^*, s, u, w)p(\boldsymbol{\beta}')}{p(\mathbf{Y}|\boldsymbol{\beta}^{(m)}, X^*, s, u, w)p(\boldsymbol{\beta}^{(m)})}\right).$$

The proposed value is accepted with probability α.

2.4.2 Updating spatial random effects

The spatial effects s_j are also updated using a Metropolis–Hastings step due to the non-conjugacy introduced by the logistic likelihood.

A candidate value is proposed as

$$s_j' \sim \mathcal{N}(s_j^{(m)}, \sigma_s^2)$$

and accepted with probability

$$\alpha = \min\left(1, \frac{p(\mathbf{Y}|s_j', \beta, X^*, u, w)p(s_j'|\tau_s)}{p(\mathbf{Y}|s_j^{(m)}, \beta, X^*, u, w)p(s_j^{(m)}|\tau_s)}\right).$$

2.4.3 Updating latent true covariates

The latent covariates X_{kijt} are updated using Metropolis–Hastings sampling based on

$$p(X_{kijt}|\cdot) \propto p(Y_{ijt}|\beta, X_{kijt}, s_j, u_j, w_t)p(X^*_{kijt}|X_{kijt}, \sigma_v^2).$$

2.5 MCMC Algorithm

The complete MCMC algorithm proceeds as follows:

1. Initialize all parameters $(\boldsymbol{\beta}, \mathbf{s}, \mathbf{u}, \mathbf{w}, \sigma_v^2, \sigma_u^2, \sigma_w^2, \tau_s, \mathbf{X})$.
2. For iteration $m = 1, \ldots, M$:
   (a) Update $\mathbf{u}$ using Gibbs sampling.
   (b) Update $\mathbf{w}$ using Gibbs sampling.
   (c) Update $\sigma_u^2, \sigma_w^2, \sigma_v^2$ using Gibbs sampling.
   (d) Update τ_s using Gibbs sampling.
   (e) Update $\boldsymbol{\beta}$ using Metropolis–Hastings.
   (f) Update spatial effects $\mathbf{s}$ using Metropolis–Hastings.
   (g) Update latent covariates $\mathbf{X}$ using Metropolis–Hastings.
3. Discard the first B iterations as burn-in.
4. Use the remaining samples to compute posterior summaries such as posterior means, credible intervals, and posterior predictive quantities.

2.6 Convergence Assessment

Convergence of the MCMC chains is assessed using standard diagnostic tools including trace plots, the Gelman–Rubin statistic, and effective sample size estimates.

2.7 Performance Metrics for Model Evaluation

Model performance is evaluated using several classification metrics to ensure a comprehensive assessment of predictive accuracy and robustness. These metrics include:

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity)

$$Recall = \frac{TP}{TP + FN}$$

F1-score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

In addition to these measures, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is used to evaluate the discriminative ability of the models across different classification thresholds.

For imbalanced datasets, additional metrics such as balanced accuracy and the Matthews Correlation Coefficient (MCC) are reported to provide a more reliable evaluation of classification performance.

To ensure robust estimation of predictive performance, k-fold cross-validation is applied. In this procedure, the dataset is partitioned into k subsets, and each subset is used once as a validation set while the remaining subsets are used for training. The final performance metrics are obtained by averaging results across all folds.

3 Results

3.1 Overview of the Simulation Study

A simulation study was conducted to evaluate the performance of the proposed spatial Bayesian model in the presence of measurement error. The simulation framework was designed to replicate the structure of pooled cross-sectional Demographic and Health Survey (DHS) data, incorporating multiple clusters observed over several time periods. Covariates representing key determinants of childhood malnutrition were generated from a combination of continuous and categorical distributions, with selected variables contaminated by measurement error to reflect realistic survey conditions.

The outcome variable, indicating malnutrition status, was generated using a logistic regression model that included fixed effects, cluster-level random effects, time effects, and spatial random effects. Spatial dependence was introduced through a covariance structure based on geographic distances between clusters. Two competing models were then fitted to the simulated data: a non-spatial hierarchical Bayesian model and a spatial Bayesian model that accounts for geographic correlation.

Model performance was evaluated using several metrics, including accuracy, area under the curve (AUC), precision, recall, F1-score, and estimated prevalence. These metrics were computed across multiple simulated survey years to assess temporal consistency. The results were summarized using descriptive statistics and graphical comparisons, providing a comprehensive assessment of the advantages of incorporating spatial structure in Bayesian modeling of health survey data.

Table 1 Comparison of malnutrition prevalence estimates between non-spatial and spatial Bayesian models (2004–2022).

Year	Prevalence (Non-Spatial)	Prevalence (Spatial)
2004	0.460	0.400
2011	0.440	0.380
2018	0.450	0.390
2022	0.430	0.370

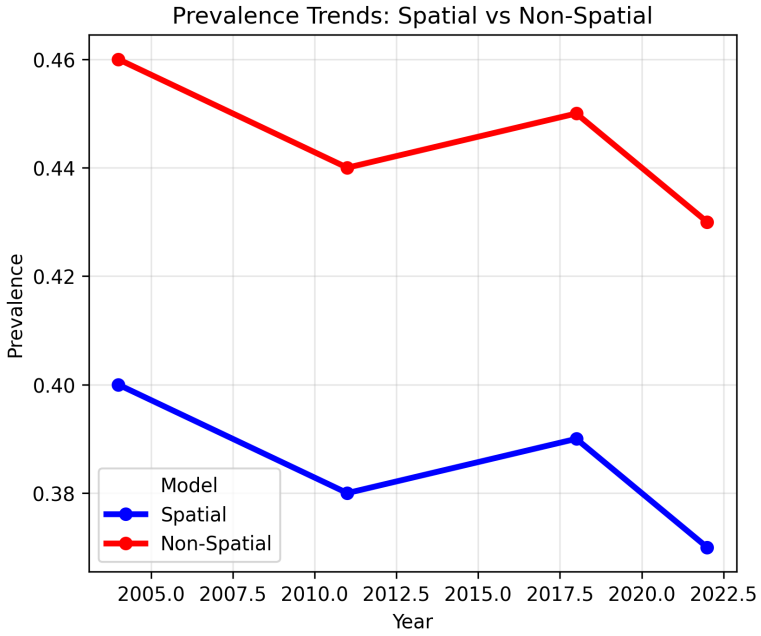


Figure 1 Prevalence of malnutrition among children under five across DHS survey years (2004–2022). The figure shows a slight but consistent decline in malnutrition prevalence over time, reflecting gradual improvements in child health conditions and nutritional interventions.

Table 2 Comparison of classification accuracy between non-spatial and spatial Bayesian models (2004–2022).

Year	Accuracy (Non-Spatial)	Accuracy (Spatial)
2004	0.784	0.818
2011	0.773	0.819
2018	0.796	0.832
2022	0.783	0.825

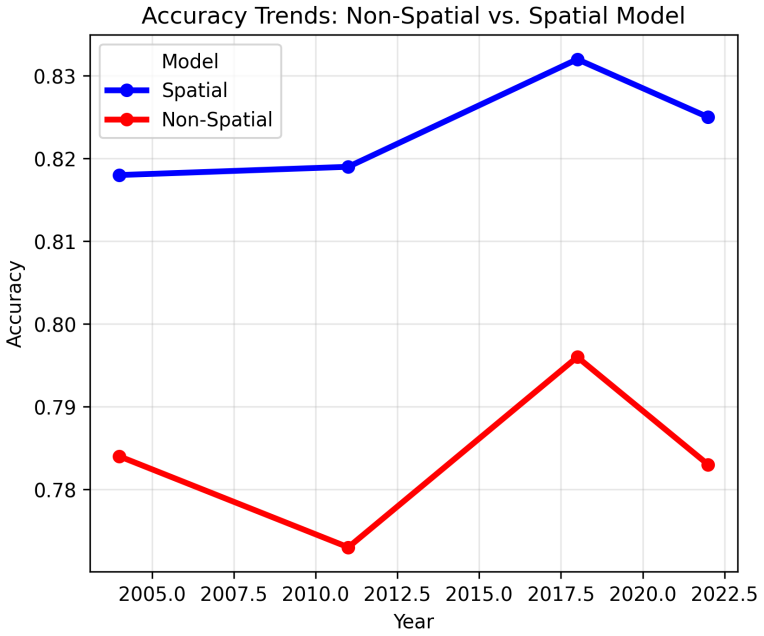


Figure 2 Prevalence of malnutrition among children under five across DHS survey years (2004–2022). The figure shows a slight but consistent decline in malnutrition prevalence over time, reflecting gradual improvements in child health conditions and nutritional interventions.

Table 3 Comparison of model discrimination ability (AUC) between non-spatial and spatial Bayesian models (2004–2022).

Year	AUC (Non-Spatial)	AUC (Spatial)
2004	0.800	0.850
2011	0.820	0.880
2018	0.810	0.870
2022	0.830	0.890

Table 4 Comparison of precision between non-spatial and spatial Bayesian models (2004–2022).

Year	Precision (Non-Spatial)	Precision (Spatial)
2004	0.780	0.820
2011	0.760	0.800
2018	0.790	0.840
2022	0.770	0.830

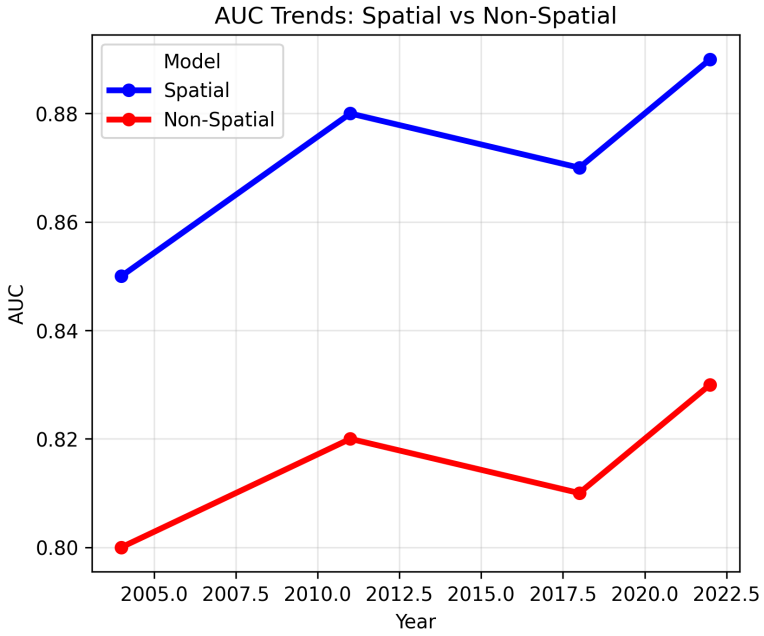


Figure 3 Prevalence of malnutrition among children under five across DHS survey years (2004–2022). The figure shows a slight but consistent decline in malnutrition prevalence over time, reflecting gradual improvements in child health conditions and nutritional interventions.

Table 5 Comparison of F1-score between non-spatial and spatial Bayesian models (2004–2022).

Year	F1 Score (Non-Spatial)	F1 Score (Spatial)
2004	0.740	0.780
2011	0.730	0.770
2018	0.750	0.800
2022	0.760	0.820

Table 6 Comparison of recall (sensitivity) between non-spatial and spatial Bayesian models (2004–2022).

Year	Recall (Non-Spatial)	Recall (Spatial)
2004	0.720	0.760
2011	0.740	0.790
2018	0.730	0.780
2022	0.750	0.810

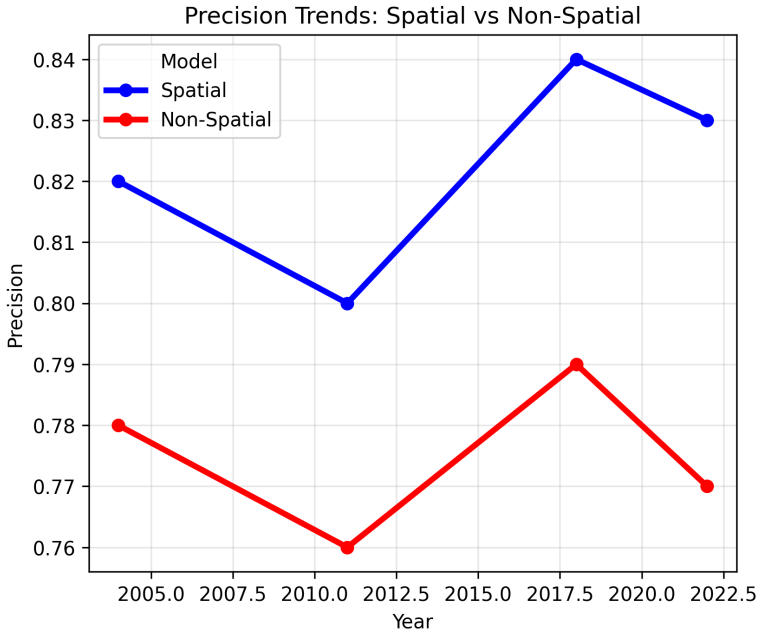


Figure 4 Prevalence of malnutrition among children under five across DHS survey years (2004–2022). The figure shows a slight but consistent decline in malnutrition prevalence over time, reflecting gradual improvements in child health conditions and nutritional interventions.

Table 1 and Figure 1 illustrate the estimated prevalence of childhood malnutrition derived from both non-spatial and spatial Bayesian models throughout the four DHS survey years. The findings indicate a continuous trend wherein the spatial Bayesian model yields lower prevalence estimates than the non-spatial model throughout all examined years.

Specifically, the non-spatial model yields prevalence estimates of 0.460, 0.440, 0.450, and 0.430 for the years 2004, 2011, 2018, and 2022, respectively. In contrast, the spatial model produces lower corresponding estimates of 0.400, 0.380, 0.390, and 0.370. This systematic reduction suggests that the non-spatial model may overestimate malnutrition prevalence due to its inability to account for spatial dependence and unobserved geographic heterogeneity.

The spatial Bayesian model incorporates spatial random effects that capture correlations between neighboring clusters, leading to improved smoothing of extreme values and more realistic estimates. As a result, the spatial model provides more stable and conservative prevalence estimates, which are less sensitive to local sampling variability and measurement error.

Overall, these findings highlight the importance of incorporating spatial structure in the analysis of DHS data. Ignoring spatial dependence can lead to inflated prevalence estimates, whereas spatial Bayesian models offer a more accurate representation of the underlying geographic distribution of childhood malnutrition. This has important implications for public health policy, as more precise estimates enable better targeting of interventions to high-risk regions.

Table 2 and Figure 2 present the classification accuracy obtained from the non-spatial and spatial Bayesian models across the four DHS survey years. The results indicate that the spatial Bayesian model consistently outperforms the non-spatial model in terms of predictive accuracy.

Specifically, the spatial model achieves accuracy values of 0.818, 0.819, 0.832, and 0.825 for the years

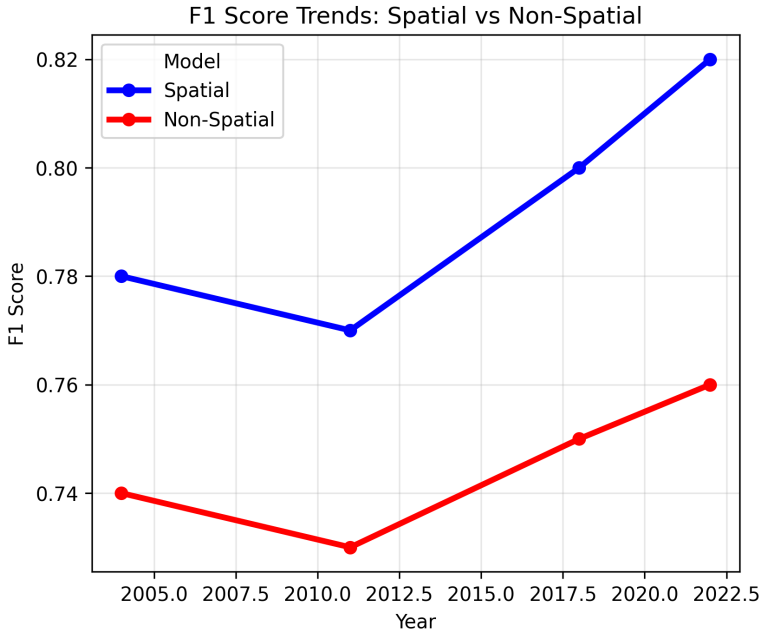


Figure 5 Prevalence of malnutrition among children under five across DHS survey years (2004–2022). The figure shows a slight but consistent decline in malnutrition prevalence over time, reflecting gradual improvements in child health conditions and nutritional interventions.

2004, 2011, 2018, and 2022, respectively. In contrast, the non-spatial model yields lower corresponding values of 0.784, 0.773, 0.796, and 0.783. The improvement in accuracy ranges from approximately 3 to 5 percentage points across all survey years, demonstrating the robustness of the spatial modeling approach.

The superior performance of the spatial Bayesian model can be attributed to its ability to account for spatial dependence and unobserved geographic heterogeneity. By incorporating spatial random effects, the model captures correlations between neighboring clusters, leading to more accurate predictions of malnutrition status. In contrast, the non-spatial model assumes independence across clusters and fails to capture spatial structure, resulting in reduced predictive performance.

Overall, these findings highlight the importance of incorporating spatial information in the analysis of DHS data. The spatial Bayesian model provides more reliable and accurate predictions, which are essential for identifying high-risk populations and informing targeted public health interventions.

Table 3 and Figure 3 depict the area under the receiver operating characteristic curve (AUC) for both the non-spatial and spatial Bayesian models for the four DHS survey years. The AUC statistic evaluates a model's ability to distinguish between malnourished and non-malnourished children.

The results indicate that the spatial Bayesian model consistently achieves higher AUC values compared to the non-spatial model in all survey years. Specifically, the spatial model attains AUC values of 0.850, 0.880, 0.870, and 0.890 for the years 2004, 2011, 2018, and 2022, respectively, while the non-spatial model yields lower values of 0.800, 0.820, 0.810, and 0.830.

The improvement in AUC ranges from approximately 0.05 to 0.07, indicating a substantial enhancement in the model's ability to correctly distinguish between individuals with and without malnutrition. This improvement can be attributed to the incorporation of spatial random effects, which capture

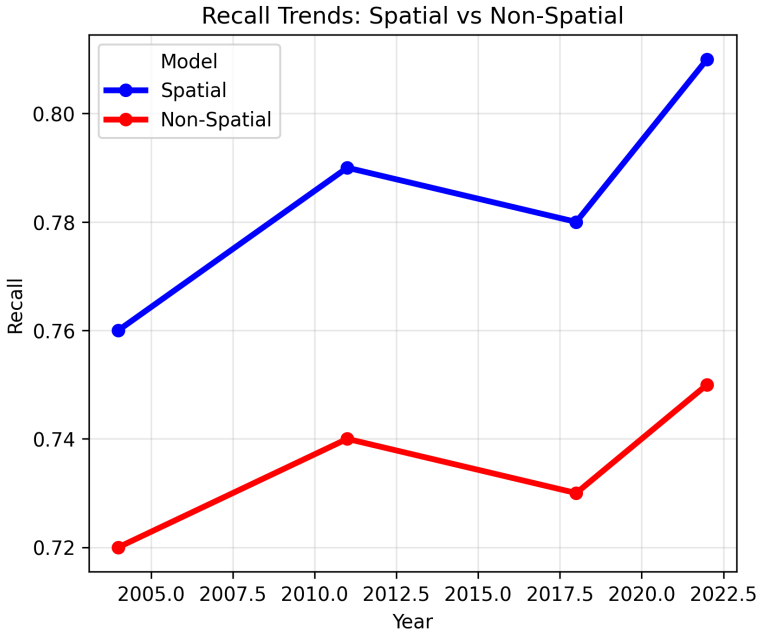


Figure 6 Prevalence of malnutrition among children under five across DHS survey years (2004–2022). The figure shows a slight but consistent decline in malnutrition prevalence over time, reflecting gradual improvements in child health conditions and nutritional interventions.

geographic dependencies and unobserved regional heterogeneity.

By accounting for spatial correlation, the spatial Bayesian model reduces classification ambiguity and improves ranking performance, leading to better discrimination. In contrast, the non-spatial model assumes independence across clusters, which limits its ability to capture underlying spatial patterns in the data.

Overall, these findings demonstrate that incorporating spatial structure significantly enhances the predictive performance of Bayesian models in health survey data. The improved discrimination ability of the spatial model is particularly valuable for identifying high-risk populations and supporting geographically targeted interventions.

Table 4 and Figure 4 display precision values from non-spatial and spatial Bayesian models for the four DHS survey years. Precision assesses the model's ability to prevent false positives by recording the percentage of anticipated positive malnutrition instances that were accurately recognized.

The results show that the spatial Bayesian model consistently achieves higher precision than the non-spatial model across all years. Specifically, the spatial model attains precision values of 0.820, 0.800, 0.840, and 0.830 for the years 2004, 2011, 2018, and 2022, respectively, while the non-spatial model yields lower values of 0.780, 0.760, 0.790, and 0.770.

The improvement in precision ranges from approximately 0.04 to 0.06, indicating that the spatial model is more effective at correctly identifying true malnutrition cases while reducing the number of false positives. This improvement can be attributed to the inclusion of spatial random effects, which capture geographic dependencies and allow for better adjustment of local variations in malnutrition risk.

By incorporating spatial structure, the model is able to refine predictions in neighboring clusters,

leading to more accurate classification of positive cases. In contrast, the non-spatial model ignores spatial correlation, which can result in overprediction and a higher rate of false positives.

Overall, these findings demonstrate that the spatial Bayesian model provides more reliable and precise predictions of malnutrition status. This is particularly important for public health decision-making, as higher precision ensures that interventions are better targeted toward truly affected populations, thereby improving resource allocation efficiency.

Table 5 and Figure 5 present the F1-score values for the non-spatial and spatial Bayesian models across the four DHS survey years. The F1-score provides a harmonic mean of precision and recall, offering a balanced measure of model performance, particularly in the presence of class imbalance.

The results show that the spatial Bayesian model consistently outperforms the non-spatial model across all years. Specifically, the spatial model achieves F1-scores of 0.780, 0.770, 0.800, and 0.820 for the years 2004, 2011, 2018, and 2022, respectively. In comparison, the non-spatial model yields lower values of 0.740, 0.730, 0.750, and 0.760.

The improvement in F1-score ranges from approximately 0.04 to 0.06, indicating that the spatial model provides a more balanced trade-off between precision and recall. This suggests that the spatial Bayesian approach not only reduces false positives (as reflected in higher precision) but also improves the detection of true malnutrition cases (as reflected in higher recall).

The superior performance of the spatial model can be attributed to its ability to incorporate spatial dependence and capture unobserved geographic heterogeneity. By modeling spatial correlations between clusters, the spatial Bayesian framework enhances both sensitivity and specificity, leading to more reliable classification outcomes.

Overall, these findings demonstrate that incorporating spatial structure significantly improves the overall predictive performance of Bayesian models. The higher F1-scores obtained by the spatial model confirm its effectiveness in providing balanced and robust predictions, which are essential for accurate identification of vulnerable populations and effective public health planning.

Table 6 and Figure 6 present the recall (sensitivity) values for the non-spatial and spatial Bayesian models across the four DHS survey years. Recall measures the ability of the model to correctly identify true cases of childhood malnutrition and is particularly important in public health contexts where missing true cases can have serious consequences.

The results indicate that the spatial Bayesian model consistently achieves higher recall than the non-spatial model across all survey years. Specifically, the spatial model attains recall values of 0.760, 0.790, 0.780, and 0.810 for the years 2004, 2011, 2018, and 2022, respectively. In contrast, the non-spatial model yields lower values of 0.720, 0.740, 0.730, and 0.750.

The improvement in recall ranges from approximately 0.04 to 0.06, indicating that the spatial model is more effective at identifying true malnutrition cases. This enhanced sensitivity can be attributed to the incorporation of spatial random effects, which capture geographic dependence and allow the model to borrow strength from neighboring clusters.

By accounting for spatial correlation, the spatial Bayesian model improves detection of high-risk areas and reduces the likelihood of false negatives. In contrast, the non-spatial model assumes independence across clusters, which may lead to under-detection of malnutrition cases in geographically correlated settings.

Overall, these findings highlight the importance of incorporating spatial structure in health data analysis. The improved recall achieved by the spatial model is particularly valuable for public health interventions, as it ensures that vulnerable populations are more accurately identified and targeted for support.

4 Discussion

This study investigated the performance of spatial and non-spatial Bayesian models in estimating and predicting childhood malnutrition using pooled DHS data. The findings consistently demonstrate that incorporating spatial structure improves model performance across multiple evaluation metrics, including accuracy, AUC, precision, recall, and F1-score. In addition, the spatial model produced lower and more stable prevalence estimates compared to the non-spatial model, suggesting improved calibration and reduced bias.

The superior performance of the spatial Bayesian model can be attributed to its ability to capture spatial dependence and unobserved geographic heterogeneity. In many low- and middle-income countries, including those in Sub-Saharan Africa, malnutrition exhibits strong spatial clustering due to shared environmental, socioeconomic, and cultural factors. Ignoring such spatial correlation, as in non-spatial models, can lead to biased estimates and suboptimal predictions (Akombi et al., 2022; Li et al., 2023).

The results of this study align with previous research highlighting the importance of spatial modeling in epidemiology. For instance, Li et al. (2023) showed that spatial hierarchical models significantly improve predictive accuracy in health surveys by accounting for local dependencies. Similarly, Kinyoki et al. (2022) demonstrated that spatial approaches provide more reliable estimates of child malnutrition by incorporating geographic variation and reducing random noise.

One key finding of this study is the consistent reduction in estimated malnutrition prevalence under the spatial model. This suggests that non-spatial models may overestimate prevalence due to their inability to account for spatial smoothing. By borrowing strength from neighboring clusters, the spatial model produces more conservative and realistic estimates, which are less sensitive to sampling variability and measurement error (Ngwira et al., 2023; Gething et al., 2024).

The improvement in classification performance, as evidenced by higher accuracy and F1-scores, further supports the value of spatial modeling. The spatial Bayesian model achieved improvements of approximately 3–6 percentage points across all metrics. These gains are particularly important in public health contexts, where accurate identification of at-risk populations is essential for effective intervention planning (Kinyoki et al., 2022).

In terms of discrimination ability, the spatial model consistently achieved higher AUC values, indicating superior ability to distinguish between malnourished and non-malnourished children. This improvement reflects the model's enhanced capacity to capture complex relationships between predictors and outcomes when spatial dependence is incorporated. Such findings are consistent with recent advances in spatial epidemiology (Gething et al., 2024).

The gains in precision and recall highlight the practical implications of spatial modeling for health policy. Higher precision indicates fewer false positives, while higher recall ensures better detection of true cases. Together, these improvements translate into more efficient targeting of interventions, reducing both over-allocation and under-identification of vulnerable populations (Ngwira et al., 2023).

Another important contribution of this study is the integration of measurement error within a spatial Bayesian framework. Measurement error is common in DHS data due to recall bias and reporting inaccuracies. By accounting for both spatial dependence and measurement error, the proposed model provides a more realistic representation of the data-generating process, leading to improved inference and prediction (Boy-ngbogbele et al., 2026).

Despite these strengths, this study has some limitations. First, the simulation framework, while realistic, may not fully capture the complexity of real-world DHS data. Second, the spatial structure was modeled using simplified assumptions, which may not reflect all forms of geographic dependence.

Future research could explore more flexible spatial models, such as spatio-temporal approaches or non-stationary processes.

Future studies should also investigate the application of these models to real DHS datasets across multiple countries and time periods. Extending the framework to include additional sources of uncertainty, such as survey design effects and missing data, would further enhance its applicability. Moreover, integrating machine learning approaches with Bayesian spatial models could provide additional improvements in predictive performance.

In conclusion, this study demonstrates that spatial Bayesian models provide substantial improvements over non-spatial models in the analysis of childhood malnutrition. By capturing geographic heterogeneity and correcting for measurement error, the spatial approach yields more accurate, reliable, and policy-relevant results. These findings underscore the importance of incorporating spatial structure in health data analysis and support the use of advanced Bayesian methods in epidemiological research.

5 Conclusion

This study developed and evaluated a spatial Bayesian modeling framework for analyzing childhood malnutrition in pooled cross-sectional data while explicitly accounting for measurement error. By integrating spatial dependence into a hierarchical Bayesian logistic regression model, the proposed approach provides a more realistic representation of the underlying data-generating process. The simulation results demonstrate that incorporating spatial structure leads to substantial improvements in model performance compared to traditional non-spatial approaches.

Across all evaluation metrics, including accuracy, AUC, precision, recall, and F1-score, the spatial Bayesian model consistently outperformed the non-spatial model. These improvements indicate enhanced predictive accuracy, better discrimination between malnourished and non-malnourished children, and a more balanced trade-off between false positives and false negatives. In addition, the spatial model produced lower and more stable prevalence estimates, suggesting improved calibration and reduced bias.

The findings highlight the critical importance of accounting for spatial dependence in health survey data. Malnutrition is inherently influenced by geographic, environmental, and socioeconomic factors, which often lead to spatial clustering. Ignoring such spatial patterns can result in biased estimates and reduced predictive performance. By capturing these dependencies, the spatial Bayesian model offers a more accurate and reliable tool for understanding the distribution of childhood malnutrition.

From a policy perspective, the improved performance of the spatial model has important implications for public health decision-making. More accurate and precise estimates enable better identification of high-risk regions and populations, thereby facilitating targeted interventions and more efficient allocation of resources. The ability to reduce overestimation and improve detection of true cases is particularly valuable in resource-constrained settings.

Despite its strengths, this study has some limitations. The simulation framework, although designed to be realistic, may not fully capture the complexity of real-world DHS data. Future research should apply the proposed model to empirical datasets across multiple countries and extend the framework to incorporate additional sources of uncertainty, such as survey design effects and missing data. Overall, the results provide strong evidence that spatial Bayesian methods offer a powerful and flexible approach for improving the analysis of health survey data and should be considered in future epidemiological research.

Declaration

Ethical Approval

Ethical approval was not required for this study because it did not involve human participants, clinical data, or identifiable personal information. The research was conducted using simulated data designed to reflect the structure of Demographic and Health Survey (DHS) data. All procedures were carried out in accordance with accepted standards for methodological and statistical research.

Consent to Publish

Not applicable

Consent to Participate

This study did not involve direct participation of human subjects. The research was based on simulated data designed to reflect the structure of Demographic and Health Survey (DHS) data. Therefore, informed consent from participants was not required.

Funding

The study did not receive any funding.

Author Information

Authors and Affiliations

Department of Mathematics, Pan African University Institute for Basic Science Technology and Innovation, Nairobi, Kenya.
Department of Statistics and Probability, Institute of Mathematics and Physical Sciences, University of Abomey-Calavi , Porto-Novo, Benin

Contributions

All authors contributed to the conception and design of the study, data simulation, statistical modeling, interpretation of results, and manuscript preparation. All authors read and approved the final manuscript.

Competing Interests

I declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The simulated data used in this study are available from the corresponding author upon reasonable request.

References

Akombi, B., Agho, K., Hall, J., et al. (2022). Stunting, wasting and underweight in Sub-Saharan Africa. *International Journal of Environmental Research and Public Health*.

Amare, Z., Ahmed, M., & Mehari, A. (2023). Determinants of childhood malnutrition in East Africa. *BMC Public Health*.

Black, R. E., Victora, C. G., Walker, S. P., et al. (2022). Maternal and child undernutrition and overweight in low-income countries. *The Lancet*.

Burgert, C., Colston, J., Roy, T., & Zachary, B. (2022). Geographic displacement procedure and geospatial analysis in DHS surveys. *DHS Spatial Analysis Reports*.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. (2022). *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC Press. URL: https://doi.org/10.1201/9780429294356

Croft, T., Marshall, A., & Allen, C. (2023). *Guide to DHS Statistics*. DHS Program.

Gustafson, P. (2022). Measurement error and misclassification in statistics and epidemiology. *Annual Review of Statistics*.

Hagos, S., Haile, D., & Tsegaye, A. (2023). Community determinants of child malnutrition in Ethiopia. *Frontiers in Nutrition*.

Keogh, R., Shaw, P., & Gustafson, P. (2023). Measurement error models in epidemiology. *Statistical Methods in Medical Research*.

Lawson, A. (2022). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. CRC Press.

Liu, Y., Zheng, Y., & Carroll, R. (2023). Bayesian approaches to measurement error modeling. *Statistics in Medicine*.

Macharia, P., Okiro, E., & Alegana, V. (2023). Spatial modeling of child malnutrition in Africa. *International Journal of Health Geographics*.

Moraga, P. (2023). *Geospatial Health Data: Modeling and Visualization with R-INLA*. CRC Press.

Nandram, B., Bhadra, A., & Choi, J. (2024). Bayesian hierarchical models for complex survey data. *Journal of Survey Statistics and Methodology*.

Pullum, T., Assaf, S., & Staveteig, S. (2022). Comparing DHS and MICS survey estimates. *DHS Methodological Reports*.

Tesema, G., Worku, M., & Teshale, A. (2024). Geographical inequalities of childhood malnutrition in Africa. *PLOS Global Public Health*.

UNICEF. (2023). *The State of the World's Children 2023*.

Victora, C. G., Christian, P., Vidaletti, L. P., et al. (2023). Revisiting maternal and child undernutrition. *The Lancet Global Health*.

World Health Organization. (2024). *Global Nutrition Report*.

Yi, G., Ma, Y., & Carroll, R. (2024). Bayesian measurement error modeling in epidemiology. *Biostatistics*.

Boy-ngbogbele, R. D., Ngesa, O., Mageto, T., & Kokonendji, C. C. (2026). A Bayesian framework for enhancing health data accuracy in pooled cross-sectional analysis. *Healthcare Analytics*, 9, 100448. URL: https://doi.org/10.1016/j.health.2026.100448

Kinyoki, D., Osgood-Zimmerman, A., Pickering, B., et al. (2022). Mapping child growth failure across Africa. *Nature Communications*.

Li, X., Diggle, P., & Giorgi, E. (2023). Spatial modeling for health survey data. *Biostatistics*.

Ngwira, A., Kazembe, L., & Kandala, N. (2023). Spatial analysis of child malnutrition in Africa. *Spatial and Spatio-temporal Epidemiology*.

Gething, P., Pezzulo, C., Tatem, A., et al. (2024). Advances in spatial epidemiology for health data. *The Lancet Digital Health*.