Bayesian Spatial Small Area Estimations of Undernutrition Using Big Data Sources

Seyifemickael Amare Yilema (samarey1981@gmail.com)
Bahir Dar University
Yegnanew A. Shiferaw
University of Johannesburg
Temesgen Zewotir
University of KwaZulu-Natal
Essey Kebede Muluneh
Bahir Dar University

Research Article

Keywords: MCMC, Fay-Herriot, Bayesian spatial models, survey, census, undernutrition

Posted Date: July 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1855365/v1

Abstract

Survey estimates play a vital role in policy decision-making at the national and regional levels. Small area estimation (SAE) for aggregated area-level data is commonly employed under the Fay–Herriot model with independently and identically distributed area effect assumption. However, the presence of a spatial influence between adjacent or surrounding regions cannot be ignored in various applications, which is not handled by this approach. Simultaneous autoregressive specifications take into account spatial connection as well as spatial correlation effects between neighboring areas. In addition, the prior parameter information is required to maximize the efficiency of direct survey estimations. The study aimed to use Bayesian spatial statistics under the Fay-Herriot model to improve direct survey estimates of undernutrition in Ethiopia. Markov Chain Monte Carlo (MCMC) algorithms were applied for implementing the Bayesian spatial FH model. The diagnostic measures are demonstrated to examine the model assumptions and reliability and validity of the generated model-based small area estimates. Bayesian spatial SAE has a lower percent coefficient of variation (CV) than the corresponding direct survey estimates of undernutrition. Therefore, model-based estimates are more precise, efficient, and adequate than the corresponding estimates for all target variables (stunting, wasting, and underweight) in almost all zones for children under five. The zonal level undernutrition estimates produced by the Bayesian spatial SAE methods provide invaluable information to policy analysts and decision-makers.

1. Introduction

Small area estimation (SAE) is typically developed to challenge unreliable and large variances of direct survey estimates due to small sample sizes (Rao and Molina 2015). Because SAEs are widely used in the public and private sectors, the demand for appropriate disaggregated level statistics from sample surveys has risen in recent decades. The SAE technique is designed primarily to generate reasonable estimates of target variables for small domains and reduced unreasonably large sampling errors. Because of the sample size issue, direct survey estimation methodologies are insufficient for generating disaggregated or small domain level estimates. As a result, SAE approaches offer a practical solution to generate credible estimates at a disaggregated level by borrowing strength from other relevant sources such as the census (Jin et al. 2018; Chandra, Salvati, and Chambers 2018; Pfeffermann and Ben-Hur 2018; Anjoy and Chandra 2019). The SAE approaches given prior information for the parameters, called Bayesian approaches, are studied by different scholars (Jin et al. 2018; Zhang and Bryant 2020; Moura and Migon 2002; Kubacki and Jedrzejczak 2016; Ghosh 1992).

The hierarchical Bayesian methods of the Fay-Herriot model were proposed by (Ghosh 1992), and the Bayesian spatial SAE framework was further studied by (Anjoy and Chandra 2019; Kubacki and Jedrzejczak 2016; Moura and Migon 2002; Anjoy and Chandra 2021; You and Zhou 2011). It has been found that Bayesian spatial approaches of SAE to the problem of small sample sizes for unplanned domains have several advantages over direct survey estimates (Anjoy and Chandra 2019). In model-based estimations, Bayesian spatial Fay-Herriot model, random area effects are significant in capturing the residual uncertainty of the small area estimates; the synthetic regression model does not explain that. Independent assumptions may not always be valid when random effects are linked. We considered the spatial random-effects models, including the popular simultaneous autoregressive models, as alternatives to the Fay-Herriot model. We implement Bayesian spatial SAE on non-informative priors for model parameters to overcome the problems of small sample sizes (Anjoy and Chandra 2021; You and Zhou 2011).

In this study, the 2016 Ethiopian demographic health survey (DHS) z scores of undernutrition (stunting, wasting, and underweight) in children under five were taken. The World Health Organization (WHO) established new child growth standards, including standardized measurements of stunting (height-for-age), wasting (weight-for-age), and underweight (weight-for-height) (De Onis 2006).

According to works of literature, an estimated 144 million and 47 million children under the age of five are stunted and wasting, respectively (De Onis 2006). Asia and Africa were home to most of the world's stunted, underweight, and wasted children under five (UNICEF, WHO, and World Bank 2020). In Ethiopia, 38%, 10%, and 24% of children under five were stunted, wasted, or underweight (CSA and ICF 2016; Tekile, Woya, and Basha 2019).

Several researchers studied undernutrition at the regional and national levels (Amare et al. 2016; Endris, Asefa, and Dube 2017; Gebre et al. 2019; Tadesse and Alemu 2015; Teshome et al. 2010; Woodruff et al. 2017; Yeshaw et al. 2020; Tekile, Woya, and Basha 2019). Only planned domains at the national and regional levels were considered in these studies. Estimates of undernutrition in the unplanned domains such as zones (the third administrative layer) in Ethiopia can be helpful for the government, policymakers, stakeholders, and so on. Due to the small sample size, the usual direct estimates obtained from the sample surveys cannot be employed at the zonal level (Rao and Molina 2015).

Over the previous two decades, periodic surveys in Ethiopia have been undertaken over broad geographical areas to meet the goals, resulting in precise direct survey estimates (national and regional levels). However, the survey overlooked smaller geographical areas such as zones, which are crucial for policy formulations. Therefore, zones are unplanned domains that have a small sample size. The zones are the third administrative layer of the Ethiopian government, where operational planning, resource allocation, and health care execution occur (Woldie, Jirra, and Azene 2011; Kitaw et al. 2012). Zonal governments serve as a bridge for woredas (districts) and higher governmental structures. As a result, addressing the problem of undernutrition and its variations among administrative zones will better understand the country's health priorities for children under the age of five (Kitaw et al. 2012; Fenta, Zewotir, and Muluneh 2021). These would significantly aid zonal health departments in making informed decisions and taking action in planning, monitoring, and evaluating health initiatives at the zonal level (Corsi et al. 2010; Chandra, Salvati, and Chambers 2015). As a result, estimates of undernutrition indicators at the zonal level are a considerable benefit for legislative bodies, policymakers, and monitors at all levels of government. This study aims to improve the direct survey estimates of the z scores of undernutrition for under-five children in Ethiopian zones using Bayesian spatial Fay-Herriot models. In this study, the Bayesian spatial FH model estimates of undernutrition are applied for the first time in Ethiopia. The approach is based on survey data from Ethiopia's 2016 DHS program, linked to the 2007 population census.

2. Methods And Materials

2.1. Data Sources and Study Variables

Ethiopia is organized into nine regions and two administrative cities, which in turn are divided into 83 zones. By integrating the features of interest from the 2016 survey with the 2007 census data, we used the Bayesian spatial Fay-Herriot model to estimate the z scores of undernutrition (stunting, wasting, and underweight). The DHS is a nationwide household survey that includes critical demographic and health indicators to monitor and assess population, health, and nutrition initiatives. DHS data is collected using standardized questionnaires that produce various data files (CSA and ICF 2016). This study combined the 2016 Ethiopian DHS and 2007 census data. The census is a big data source since it is a total enumerations of the population and housing census data (Ghosh, Nangia, and Kim 1996; Maples 2017; Checchi et al. 2022; Marchetti et al. 2015). These datasets are collected by Ethiopia's central statistical agency (CSA 2007; CSA and ICF 2016).

2.1.1 Spatial data

We matched all household characteristics to the global positioning system (GPS) point data for each sampled urban-rural cluster household. For confidentiality reasons, the GPS urban/rural locations have been masked (Burgert et al. 2013). Urban household clusters were displaced up to 2 km and rural residence clusters up to 5 km in small-area administrative units, with 1% of rural clusters displaced up to 10 km (Burgert et al. 2013).

2.1.2 Sampling design

In the 2016 Ethiopian DHS sample, a two-stage stratifying sampling method was used to reach all nine regions, two administration cities, and urban and rural areas. A total of 21 sampling strata were identified. In two stages, samples of enumeration areas were selected independently from each stratum. At each lower administrative level, implicit stratification and proportional allocation were achieved (CSA and ICF 2016)

645 enumeration areas were chosen randomly in the first stage in each stratum with a probability proportional to the size of the enumeration area. There were 202 enumeration areas planned for urban and 443 enumeration areas for rural areas. In the second stage, an equal probability systematic selection was used to choose 28 households per cluster from newly formed household lists. Height and weight measurements for z scores of undernutrition were obtained from children aged 0 to 59 months (CSA and ICF 2016).

2.2. Small Area Estimation under Fay Herriot Model

Let be the number of small areas (zones in our cases), y_i be the direct survey estimates of the population parameter θ_i for the area . Moreover, let $\mathbf{x}_i$ be a $p \times 1$ vector of auxiliary variables for the area often taken from the census records (Rao and Molina 2015; Yilema et al. 2022b). The general Fay-Herriot model, which is proposed by (Fay and Herriot 1979), is given as follows

$$y_i = \theta_i + \dot{o}_i \quad \text{...} \quad (1)$$

$$\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \nu_i \quad \text{..} \quad (2)$$

The first model (1) is the sampling model, which accounts for the sampling variability of the direct survey estimates y_i of the population parameters θ_i. The second model (2) is the linking model, which links the population parameter θ_i to the known auxiliary variables. The Fay-Herriot model is a small-area estimation model at the area level that combines the two models described below.

$$y_i = \mathbf{x}_i^t \beta + \nu_i + \epsilon_i \text{..} (3)$$

Where β are the regression coefficients of the auxiliary variables, ν_i being the area of random effects that are independently and identically distributed ($i.\ i.\ d$) with $E(\nu_i) = 0$ and $var(\nu_i) = \sigma_\nu^2 I_m$, and the sampling variance ϵ_i are independently distributed with mean zero and variance $\sigma_{\epsilon_i}^2$, $\epsilon_i \sim N(0, \sigma_{\epsilon_i}^2)$. The two random errors ϵ_i and ν_i are independent of each other within and across the areas (Fay and Herriot 1979; Rao and Molina 2015; You and Zhou 2011; Yilema et al. 2022a). The sampling $\sigma_{\epsilon_i}^2$($i = 1, \cdots, m$) is known but practically calculated from the survey data. The matrix notation of the Fay-Herriot model is given by

$$\mathbf{y} = \theta + \epsilon = \mathbf{x}\beta + \nu + \epsilon \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (4)$$

Where $\mathbf{y} = (y_1, \cdots, \theta_m)^T$ is the $m \times 1$ vector of direct survey estimates, $\theta = (\theta_1, \cdots, \theta_m)^T$ is the $m \times 1$ vector of population parameters, $\mathbf{x} = (x_1, \cdots, x_m)^T$ is the $m \times p$ matrix of auxiliary variables, $\nu = (\nu_1, \cdots, \nu_m)^T$ is the $m \times 1$ vector of random area effects with mean zero and variance $\sigma_\nu^2 I_m$, and $\epsilon = (\epsilon_1, \cdots, \epsilon_m)$ is $m \times 1$ vector of sampling errors with $\epsilon \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$, where $\Sigma_\epsilon =$ $diag\{\sigma_{\epsilon_i}^2, 1 \leqslant i \leqslant m\}$ is the design variance. Therefore, the variance-covariance matrix of the vector $\mathbf{y}$ is $Var(y) = \mathbf{V} = \sigma_\nu^2 \mathbf{I}_m + \boldsymbol{\Sigma}_\epsilon$, I_m is $m \times m$ the identity matrix.

2.3. Bayesian Spatial Small Area Estimation

The underlying assumption in the Fay-Herriot model is that direct survey estimates from diverse local areas are uncorrelated (Anjoy and Chandra 2021). In practice, however, the boundaries that constitute a small region are often arbitrary, and there appears to be no convincing reason why neighboring areas should not be correlated. As a result, it is common to presume that the effects of nearby small areas, as determined by a contiguity requirement, are correlated (Chandra, Salvati, and Chambers 2015; Pratesi and Salvati 2008; Petrucci et al. 2006). Small area modeling enhances model accuracy by including spatial information between surrounding areas. As a result, the so-called SAR error process is frequently used to combine spatial data into a linking model with spatial dependence in error structure. Let us define the random area impact that we want to achieve.

$\nu = (\mathbf{I}_m - \rho\mathbf{W})^{-1}\mu$..(5)

Here $\mu = (\mu_1, \ldots, \mu_m)^T$ is a vector with $E(\mu) = \mathbf{0}$ and covariance matrix $\sigma_\mu^2\mathbf{I}_m$, which $\mathbf{I}_m$ indicates the $m \times m$ identity matrix and σ_μ^2 is an unknown parameter. With the assumption that the matrix $(\mathbf{I}_m - \rho\mathbf{W})$ is non-singular and {\mathbf{\nu }} is mean E({\mathbf{\nu }})={\mathbf{0}} and variance var({\mathbf{\nu }})= **G** = \sigma _{\nu }^{2}{[({{\mathbf{I}}_m} - \rho {{\mathbf{W}}^T})({{\mathbf{I}}_m} - \rho {{\mathbf{W}}^T})]^{ - 1}}. The proximity matrix **W** is defined in row standardized form, i.e., **W** is row stochastic in \rho \in (- 1,1) the SAR parameter (Petrucci et al. 2006; Yilema et al. 2022c). \rhois the spatial autoregressive coefficient, which quantifies the intensity of the geographical link, and {\mathbf{W}} is the proximity or contiguity matrix that specifies the relation of random effects from surrounding areas. The simplest way to define spatial interaction between areas neighboring domains is to use a contiguity matrix **W**. The choice of proximity matrix has been employed in a number of studies (Chandra 2013; Anjoy and Chandra 2019).

The following two-stage models are given the Bayesian spatial FH model version of the general FH model (Fay and Herriot 1979).

Sampling model: {\mathbf{y}}|{\mathbf{\theta }},{\mathbf{\beta }},\rho ,\sigma _{\nu }^{2}\sim N({\mathbf{\theta }},{{\mathbf{\Sigma }}_\epsilon }) ……………………………………………. (6)

Linking model: {\mathbf{\theta }}|{\mathbf{\beta }},\rho ,\sigma _{\nu }^{2}\sim N({\mathbf{X\beta }}, {\mathbf{G}})……………………………………………… (7)

$\pi(\rho, \sigma_v^2) \sim uniform(-1,1)$ .... (8)

Where **β** is the -component regression coefficient vector \sigma _{\nu }^{2} is the model-based random component variance and \rhois the spatial autocorrelation that captures the strength of spatial dependence. The matrix notation of the Bayesian spatial Fay-Herriot model is given in analogs to the Fay Herriot model in model (4) as

{\mathbf{y}}={\mathbf{x\beta }}+{({{\mathbf{I}}_m} - \rho {\mathbf{W}})^{ - 1}}{\mathbf{\nu }}+\epsilon \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots (9)

Where the notations ${\mathbf{y}},{\mathbf{x}},{\mathbf{\beta }},{\mathbf{\nu }},\epsilon$ are the same as the notation in model (4), ρ is the simultaneous autoregression parameters, and ${{\mathbf{I}}_{\mathbf{m}}}$ are the $m \times m$ identity matrix.

2.3.1 Parameter estimations using MCMC

To estimate the characteristics parameter of interest (undernutrition), we apply a hierarchical Bayes (HB) approach using the Gibbs sampling method for the Bayesian spatial FH model. Accordingly, the potential Bayesian analog of the FH model can be suitably considered rather than its frequentist version (Gelman et al. 2006). This article mainly used the HB spatial method to employ the Gibbs sampling method. In the HB spatial method, together with the prior distribution of the parameters, prior of the hyper-parameters (model parameters) are also specified. Inferences are made from the posterior distributions. In estimating a parameter, posterior means are computed, and posterior variance is used to measure errors or uncertainties. HB spatial method can effectively deal with complex small area models using the Monte Carlo Markov Chain (MCMC) method, which overcomes the computational difficulties of high-dimensional integrations of posterior densities (Anjoy and Chandra 2021; Gelman et al. 2021; Anjoy and Chandra 2019; Moura and Migon 2002).

The priors are also another crucial component in Bayesian analysis. The uniform prior (8) on the model parameters is a popular non-informative prior (Chung and Datta 2020; Shi 2018). The priors are chosen based on works of literature as standard normal distributions for ${\mathbf{\beta }}$ uniform distribution for $\rho$and $\sigma _{\nu }^{2}$ (You and Chapman 2006; Shi 2018; Arima et al. 2017). Bayesian spatial small area predictors of ${\mathbf{\theta }}$and ${\mathbf{\beta }}$ can provide good starting values for the MCMC procedure (Shi 2018; Gelman et al. 2021). In this work, the coefficients of the synthetic regression model are used as the initial values ${\mathbf{\beta }}$, where the response vector is the direct survey estimates regressed on the design matrix in the linking model (Gelman et al. 2021).

3. Results

In this model, 41 auxiliary variables (proportions) taken from the 2007 population census of Ethiopia were considered (Ghosh, Nangia, and Kim 1996; Marchetti et al. 2015; Checchi et al. 2022). However, suitable variables were chosen using stepwise regression analysis. For z scores of stunting, age groups of 15–24 years, children aged 4–5 years, non-disabled parents, other marital status (divorced or widowed), unemployed, illiterate, and no death of a family's females were chosen. The z scores of wasting were age groups 15–24 years, children aged 4–5 years, other marital status (divorced or widowed), children aged 2–3 years, government employment, and death of just one daughter in a household. For z scores of underweight, the number of children under the age of five, other marital status (divorced or widowed), married, employer, government employee, and better water facility were chosen.

3.1. Model Diagnostics

Even though the diagnostic measures in the Fay-Herriot model are not advanced, two diagnostic criteria are commonly used: the model diagnostic and the small area estimates diagnostic (Anjoy and Chandra 2021; Anjoy, Chandra, and Basak 2019). The first diagnostic method is used to prove the model assumptions. In contrast, the second diagnostic method is used to validate the reliability of model-based small area estimations of undernutrition produced by the Bayesian spatial Fay-Herriot model. Random area effects should have an independent and identical normal distribution with mean zero and constant variance in the commonly used Fay-Herriot model. However, in spatial models, random area effects (i.e., m-zonal effects) are randomly correlated.

The bias scatter plots of undernutrition indicators for the direct survey and model-based estimates are shown in Fig. 1. Bias diagnostics are used to determine if model-based estimations of undernutrition are more or less extreme than direct survey undernutrition estimates. Plots showing direct survey estimates on the vertical axis against Bayesian spatial small area estimates on the horizontal axis, with the fitted regression line safe for divergence and tested for intercept =0 and slope =1 (Brown et al. 2001; Chandra, Salvati, and Chambers 2015). The results in Fig. 1 reveal that the Bayesian spatial small area estimates are not as extreme as the direct survey estimates, implying that the Bayesian spatial Fay-Herriot model approach reduced more extreme values toward the average. The regression line is linear and parallel to the identity line (y=x) for actual values. As a result, if the small area estimates are near-real values, the regression of the direct survey estimates on the Bayesian spatial Fay-Herriot model estimates is essentially identical. For most zones, Bayesian spatial small area estimates of undernutrition are aligned with the line, showing that they are designed unbiased (Guha and Chandra 2021b; 2021a).

The direct survey estimates and Bayesian spatial small area estimates of undernutrition under the Bayesian spatial Fay-Herriot model are shown in Fig. 2. We can observe that the model-based estimates are slightly smoother than the direct survey estimates. This is because of the auxiliary variables linked to the direct survey estimates using the Bayesian spatial Fay-Herriot model (Molina and Marhuenda 2015). According to the outcomes of both estimations, the direct survey estimates are compatible with the Bayesian spatial small area estimates of undernutrition under Bayesian spatial Fay-Herriot models (Molina and Marhuenda 2015).

3.2. Improvements in Direct Survey Estimates

Table 1 shows the Bayesian spatial SAE and direct survey estimates of undernutrition for children under five. The Bayesian spatial Fay-Herriot model technique is used in this study to produce Bayesian spatial small area estimates of undernutrition for several zones in Ethiopia. Direct survey undernutrition estimates are produced to compare with Bayesian spatial small area estimates of undernutrition.

The coefficient of variation (percent CV) is one of the key criteria for determining which model has the best performance and the most reliable estimates. Estimates with a lower percent CV are more precise and reliable than those with a higher percent CV. However, estimates with a higher percent CV are unreliable (Hossain et al. 2020; Chandra, Aditya, and Sud 2018). The percent CV of direct survey estimates for stunting, wasting, and underweight varies from 3.26 to 22.30, 3.40 to 24.32, and 3.84 to 34.02, respectively. The percent CV of Bayesian spatial small area estimates for stunting, wasting, and underweight varied from 0.79 to 6.02, 0.74 to 4.47, and 0.84 to 5.56, respectively, according to the summary data in Table 1.

The percent CV of Bayesian spatial SAE techniques is often lower than those of direct survey estimates of undernutrition. As a result, Bayesian spatial small area estimation strategies increase the precision and reliability of direct survey estimates (Brakel and Boonstra 2021; Guha and Chandra 2021a; Rahman, Moloy, and Salan 2021; Guha and Chandra 2021b). Furthermore, the findings imply that estimating undernutrition with a Bayesian spatial Fay-Herriot model significantly improves the efficiency of the model-based estimates over direct survey estimates.

Table 1
Percentage CV of Bayesian spatial SAE and direct survey estimates of undernutrition

Statistic	**stunting**		**Wasting**		**Underweight**	
	Direct estimate	**Bayesian estimate**	**Direct estimate**	**Bayes estimate**	**Direct estimate**	**Bayes estimate**
Minimum	3.260	0.79	3.84	0.74	3.40	0.840
1st quartile	5.295	1.275	5.665	1.220	6.71	1.445
Median	7.300	1.820	7.900	1.710	9.00	1.900
Mean	8.971	2.134	9.034	1.905	10.68	2.254
3rd quartile	11.110	2.485	10.770	2.250	12.79	2.725
Maximum	22.300	6.020	24.320	4.470	34.02	5.850

Figure 3 shows the percent CV of undernutrition indicators for both model-based and direct estimates. The percent CV examines the precision of the model-based Bayesian spatial small area estimates. The percent CV of direct survey estimates in (Fig. 3) is substantially larger than the comparable model-based Bayesian spatial small area estimates under the Bayesian spatial Fay-Herriot model. Estimates having a lower percent CV are often thought to be more precise and reliable (Chandra, Aditya, and Sud 2018; Hossain et al. 2020; Chandra 2013). The maximum percent CV of direct survey estimates is 22, 24 and 34 for stunting, wasting and underweight, respectively. At the same time, Bayesian spatial small area estimates for stunting, wasting, and underweight have maximum values of about 6, 4.5, and 5.85, respectively.

Table 2 shows the efficiency benefit of Bayesian spatial small area estimates in the percent CV over direct survey estimates of undernutrition. The efficiency improvement in percent CV was recorded with minimal values of 68.73, 73.84, and 69.11 for stunting, wasting, and underweight, respectively. At the same time, the maximum efficiency gains in percent CV were 82.38, 84.65 and 84.55 for stunting wasting and underweight, respectively. These findings reveal that Bayesian spatial Fay-Herriot models significantly improve direct survey estimates.

Table 2
Summary results of efficiency gain in the percent CV of Bayesian spatial SAE over direct survey estimates of undernutrition

Statistic	stunting	Wasting	underweight
Minimum	68.73	73.84	69.11
1st quartile	75.47	77.89	77.77
Median	76.29	78.61	78.62
Mean	76.13	78.68	78.58
3rd quartile	77.06	79.30	79.70
Maximum	82.38	84.65	84.55

The interpretations of the root MSE of undernutrition are similar to the CV's couterparts. Figure 4 displayed the root MSE of undernutrition with comparisons of direct and Bayesian spatial FH model-based estimates. The large standard errors of the direct survey estimates are reduced by using the model-based Bayesian spatial small area estimates. Thus, one of the primary motivations of small area models is achieved.

Figure 5 shows the map of Bayesian spatial small area estimates of undernutrition using the Fay-Herriot model. The map shows undernutrition estimates for all stunting, wasting, and underweight throughout Ethiopian zones. In particular, the maps show the unequal distribution for all three undernutrition indicators at the zonal level in Ethiopia.

4. Discussion And Conclusion

This article provides the zonal level estimates of undernutrition in Ethiopia using the 2016 DHS survey and the 2007 population census (CSA and ICF 2016). Bayesian spatial SAE model was applied to improve the direct survey estimates of undernutrition by combining the two datasets. The demand for small area estimations is increasing nowadays (Rao and Molina 2015). There is, however, a relatively little study in Ethiopia on SAEs using the Bayesian spatial SAE approach. As a result, this work intended to increase direct survey estimates of undernutrition by using Bayesian spatial SAE approaches for unplanned areas (zones in our instance).

Different kinds of literature examine the model diagnostics of SAE models (Anjoy and Chandra 2021; Chandra 2013; Hossain et al. 2020; Chandra, Aditya, and Sud 2018). This study used a similar model diagnostics approach for the spatial SAE model in a Bayesian framework. Therefore, the model assumptions and Bayesian spatial small area estimates are valid (Hossain et al. 2020; Chandra, Aditya, and Sud 2018; Anjoy and Chandra 2021; 2019; Petrucci et al. 2006; Brown et al. 2001).

The direct survey estimates have a higher percent CV than Bayesian spatial small area estimates. For stunting, wasting, and underweight, the model-based Bayesian spatial small area estimates reduced the variability of the direct survey estimates from 22.3 percent to 6.02 percent, 24.32 percent to 4.47 percent, and 34.02 percent to 5.85 percent, respectively. The spatial random effects, prior information provided, and auxiliary information from the census data are all explored in the literature (Trevisani and Torelli 2017; Chandra 2013; Fay and Herriot 1979; Trevisani and Torelli 2007) and are responsible for the reductions in large standard errors. This is one of the benefits of using Bayesian spatial SAE to address the issue of small sample sizes for estimating the extent of unplanned domains (zones). Bayesian spatial small area models provide a more comprehensive set of tools to handle complex and more realistic models and get reliable measures of variability (Iriondo-perez, Sukasih, and Harter 2018; Anjoy and Chandra 2021; 2019; Zhang and Bryant 2020; Kubacki and Jedrzejczak 2016; Moura and Migon 2002).

The precision of direct survey estimates is improved using Bayesian spatial Fay-Herriot models. Compared to direct survey estimates, the efficiency increase in percent CV due to the use of Bayesian spatial small area estimate is relatively substantial. In terms of the percent CV, the Bayesian spatial small area approaches had a mean efficiency gain of 76.13 percent, 78.68 percent, and 78.58 percent over direct estimates for stunting, wasting, and underweight, respectively. As a result, compared to direct survey estimates of undernutrition for children under five, the Bayesian spatial SAE methods deliver the most precise and suitable estimates.

The maps depict the spatial distribution of all undernutrition indicators at the country's zonal level, based on the estimates provided. The most stunted children are clustered in the country's North East, North West, and southern regions. On the other hand, most wasted children are concentrated in the country's North East, Eastern, and South East regions. Similarly, most of Ethiopia's underweight children live in the north, North East, Western, and Southern regions.

We can overcome small sample size concerns by obtaining cost-effective and reliable disaggregated level estimates from existing survey data by combining survey data and adopting SAE techniques under the Bayesian spatial Fay-Herriot model. This analysis also illustrates the benefits of using Bayesian spatial small area estimations over the direct survey estimates of undernutrition. A disaggregated level estimates of undernutrition for children under five are extremely useful for policy development and resource allocation to Ethiopian zones and the Ethiopian governments at large. Different stakeholders, government ministries, and non-governmental organizations can use the estimates and spatial maps provided by the model-based Bayesian spatial Fay-Herriot model estimates for policy planning and decision making. In addition, the results can provide for international organizations in their policy planning to formulate effective action plans to achieve the 2030 sustainable development goals.

Abbreviations

CV: Coefficient of Variation

DHS: Demographic Health Survey

GPS: Global Positioning System

HB: Bayesian Hierarchical

MCMC: Monte Carlo Markov Chain

MSE: Mean Square Error

SAE: Small Area Estimation

Declarations

Disclosure statement

There is no competing interest to declare.

Ethics approval and consent to participate

The research committee of Bahir Dar University Department of Statistics ethically approved this research. The authors used secondary data obtained from the Ethiopian Central Statistical Agency, and all methods were conducted according to the relevant guidelines. At the time of data collection, each subject and/or their legal guardian provided informed consent.

Consent for publication

Not applicable

Availability of data and materials

The 2016 DHS data for this study was received following a registration request from the DHS program website https://www.dhsprogram.com. The GPS enumeration area shapefiles were collected similarly from https://www.dhsprogram.com. Shapefiles for Ethiopian zonal administrative boundaries are also accessible on the website https://africaopendata.org. For research purposes, the Ethiopian Central Statistical Agency gives access to a 10% sample of the 2007 census.

Competing interests

The authors declare that they have no competing interests

Funding

This research was not supported by any grant from the public, commercial, or non-profit funding agencies.

Authors' contributions

SAY was involved in this study from data management, data analysis, drafting, and revising the final manuscript. YAS, TZ and EKM contributed to the conception, design, and interpretation of data, as well as to manuscript reviews and revisions. All authors have read and approved the manuscript.

Acknowledgements

We wish to express our gratitude to the DHS program and Ethiopian central statistical agency for providing us to access the data. In addition, we would like to thank Bahir Dar University and Debre Tabor University.

References

1. Amare, Desalegne, Ayenew Negesse, Baye Tsegaye, Birtukan Assefa, and Birehanu Ayenie. 2016. “Prevalence of Undernutrition and Its Associated Factors among Children below Five Years of Age in Bure Town, West Gojjam Zone, Amhara National Regional State, Northwest Ethiopia.” *Advances in Public Health* 2016: 8. https://doi.org/10.1155/2016/7145708.
2. Anjoy, Priyanka, and Hukum Chandra. 2019. “Hierarchical Bayes Aggregated Level Spatial Model for Crop Yield Estimation.” Journal of the Indian Society of Agricultural Statistics 73 (2): 143–52. .
3. ———. 2021. “Hierarchical Bayes Estimation of Small Area Means under a Spatial Nonstationary Fay–Herriot Model.” *Communications in Statistics: Simulation and Computation*, 1–19. https://doi.org/10.1080/03610918.2021.1926501.
4. Anjoy, Priyanka, Hukum Chandra, and Pradip Basak. 2019. “Estimation of Disaggregate-Level Poverty Incidence in Odisha Under Area-Level Hierarchical Bayes Small Area Model.” Social Indicators Research 144 (1): 251–73. https://doi.org/10.1007/s11205-018-2050-9.
5. Arima, Serena, William R. Bell, Gauri S. Datta, Carolina Franco, and Brunero Liseo. 2017. “Multivariate Fay–Herriot Bayesian Estimation of Small Area Means under Functional Measurement Error.” Journal of the Royal Statistical Society. Series A: Statistics in Society 180 (4): 1191–1209. https://doi.org/10.1111/rssa.12321.
6. Brakel, Jan A Van Den, and Harm-jan Boonstra. 2021. “Estimation of Domain Discontinuities Using Hierarchical Bayesian Fay-Herriot Models.” Survey Methodology 47 (1): 151–89.
7. Brown, Gary, Ray Chambers, Patrick Heady, and Dick Heasman. 2001. “Evaluation of Small Area Estimation Methods - An Application to Unemployement Estimats from the UK LFS.” In *Proceedings of Statistics Canada Symposium*, 1–10. https://doi.org/10.1201/9780203166314.ch1.
8. Burgert, Clara R, Josh Colston, Thea Roy, and Blake Zachary. 2013. “Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys. DHS

Spatial Analysis Reports No. 7. Calverton, Maryland, USA: ICF International.”

9. Chandra, Hukum. 2013. “Exploring Spatial Dependence in Area-Level Random Effect Model for Disaggregate-Level Crop Yield Estimation.” Journal of Applied Statistics 40 (4): 823–42. https://doi.org/10.1080/02664763.2012.756858.
10. Chandra, Hukum, Kaustav Aditya, and U. C Sud. 2018. “Localised Estimates and Spatial Mapping of Poverty Incidence in the State of Bihar in India—An Application of Small Area Estimation Techniques.” PLoS ONE 13 (6): 1–14. https://doi.org/10.1371/journal.pone.0198502.
11. Chandra, Hukum, Nicola Salvati, and Ray Chambers. 2015. “A Spatially Nonstationary Fay-Herriot Model for Small Area Estimation.” Journal of Survey Statistics and Methodology 3 (2): 109–35. https://doi.org/10.1093/jssam/smu026.
12. ———. 2018. “Small Area Estimation under a Spatially Non-Linear Model.” *Computational Statistics and Data Analysis* 126. https://doi.org/10.1016/j.csda.2018.04.002.
13. Checchi, Francesco, Adrienne Testa, Amy Gimma, Emilie Koum-Besson, and Abdihamid Warsame. 2022. “A Method for Small-Area Estimation of Population Mortality in Settings Affected by Crises.” Population Health Metrics 20 (1). https://doi.org/10.1186/s12963-022-00283-6.
14. Chung, Hee Cheol, and Gauri Sankar Datta. 2020. “Bayesian Hierarchical Spatial Models for Small Area Estimation.”
15. Corsi, Daniel J., Clara K. Chow, Scott A. Lear, M. Omar Rahman, S. V. Subramanian, and Koon K. Teo. 2010. “Shared Environments: A Multilevel Analysis of Community Context and Child Nutritional Status in Bangladesh.” Public Health Nutrition 14 (6): 951–59. https://doi.org/10.1017/S1368980010003356.
16. CSA. 2007. “Population and Housing Census of Ethiopia.” Adds Ababa.
17. CSA, and ICF. 2016. “Ethiopia Demographic Health Survey. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF.”
18. Endris, Neima, Henok Asefa, and Lamessa Dube. 2017. “Prevalence of Malnutrition and Associated Factors among Children in Rural Ethiopia.” BioMed Research International 2017: 6. https://doi.org/10.1155/2017/6587853.
19. Fay, RE, and RA Herriot. 1979. “Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data.” Journal of the American Statistical Association 74 (366): 269–77. https://doi.org/10.1080/01621459.1979.10482505.
20. Fenta, Haile Mekonnen, Temesgen Zewotir, and Essey Kebede Muluneh. 2021. “Spatial Data Analysis of Malnutrition among Children Under-Five Years in Ethiopia.” BMC Medical Research Methodology 21 (1): 1–13. https://doi.org/10.1186/s12874-021-01391-x.
21. Gebre, Abel, P Surender Reddy, Afework Mulugeta, Yayo Sedik, and Molla Kahssay. 2019. “Prevalence of Malnutrition and Associated Factors among Under-Five Children in Pastoral Communities of Afar Regional State, Northeast Ethiopia: A Community-Based Cross-Sectional Study.” *Journal of Nutrition and Metabolism* 2019: 13. https://doi.org/10.1155/2019/9187609.

22. Gelman, Andrew, J. B. Carlin, H. S. Stern, and D. B. Rubin. 2006. *Bayesian Data Analysis*. CRC Press Company.
23. Gelman, Andrew, John B Carlin, Hal S Stern, Rubin B Donald, David B Dunson, and Aki Vehtari. 2021. *Bayesian Data Analysis*. 3rd ed.
24. Ghosh, Malay. 1992. "Hierarchical and Empirical Bayes Multivariate Estimation." Institute of Mathematical Statistics 17: 151–77.
25. Ghosh, Malay, Narinder Nangia, and Dal Ho Kim. 1996. "Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach." Journal of the American Statistical Association 91 (436). https://doi.org/10.1080/01621459.1996.10476710.
26. Guha, Saurav, and Hukum Chandra. 2021a. "Measuring and Mapping Disaggregate Level Disparities in Food Consumption and Nutritional Status via Multivariate Small Area Modelling." Social Indicators Research 154 (2): 623–46. https://doi.org/10.1007/s11205-020-02573-8.
27. ———. 2021b. "Measuring Disaggregate Level Food Insecurity via Multivariate Small Area Modelling: Evidence from Rural Districts of Uttar Pradesh, India." *Food Security* 13 (3): 597–615. https://doi.org/10.1007/s12571-021-01143-1.
28. Hossain, Md Jamal, Sumonkanti Das, Hukum Chandra, and Mohammad Amirul Islam. 2020. "Disaggregate Level Estimates and Spatial Mapping of Food Insecurity in Bangladesh by Linking Survey and Census Data." PLoS ONE 15 (4): 1–16. https://doi.org/10.1371/journal.pone.0230906.
29. Iriondo-perez, Jeniffer, Amang Sukasih, and Rachel Harter. 2018. "Comparing Direct Survey and Small Area Estimates of Health Care Coverage in New York," 2018.
30. Jin, Ick Hoon, Fang Liu, Evercita C. Eugenio, Kisung You, and Ying Yuan. 2018. "Bayesian Hierarchical Spatial Model for Small Area Estimation with Non-Ignorable Nonresponses and Its Applications to the NHANES Dental Caries Assessments." The Annals of Applied Statistics, 1–20. http://arxiv.org/abs/1810.05297.
31. Kitaw, Yayehyirad, Gebre-Emanuel Teka, Hailu Meche, Hailemariam Damen, and Misganaw Fentahun. 2012. *The Evolution of Public Health in Ethiopia*. 2nd ed.
32. Kubacki, Jan, and Alina Jedrzejczak. 2016. "Small Area Estimation of Income under Spatial Sar Model." Statistics in Transition 17 (3): 365–90. https://doi.org/10.21307/stattrans-2016-028.
33. Maples, Jerry J. 2017. "Improving Small Area Estimates of Disability: Combining the American Community Survey with the Survey of Income and Program Participation." Journal of the Royal Statistical Society. Series A: Statistics in Society 180 (4). https://doi.org/10.1111/rssa.12310.
34. Marchetti, Stefano, Caterina Giusti, Monica Pratesi, Nicola Salvati, Fosca Giannotti, Dino Pedreschi, Salvatore Rinzivillo, Luca Pappalardo, and Lorenzo Gabrielli. 2015. "Small Area Model-Based Estimators Using Big Data Sources." Journal of Official Statistics 31 (2): 263–81. https://doi.org/10.1515/JOS-2015-0017.
35. Molina, Isabel, and Yolanda Marhuenda. 2015. "Sae: An R Package for Small Area Estimation." R Journal 7 (1): 81. https://doi.org/10.32614/rj-2015-007.

36. Moura, Fernando A.S., and H. S. Migon. 2002. “Bayesian Spatial Models for Small Area Estimation of Proportions.” Statistical Modeling 2 (3): 183–201. https://doi.org/10.1191/1471082x02st032oa.
37. Onis, Mercedes De. 2006. “WHO Child Growth Standards.” *Geneva: WHO*.
38. Petrucci, Alessandra, Nicola Salvati, Alessandra Petrucci, and Nicola Salvati. 2006. “Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment.” Agricultural, Biological, and Environmental Statistics 11 (2): 169–82. https://doi.org/10.1198/108571106X110531.
39. Pfeffermann, Danny, and Dano Ben-Hur. 2018. “Estimation of Randomisation Mean Square Error in Small Area Estimation.” International Statistical Review 87: 1–19. https://doi.org/doi:10.1111/insr.12289.
40. Pratesi, Monica, and Nicola Salvati. 2008. “Small Area Estimation: The EBLUP Estimator Based on Spatially Correlated Random Area Effects.” Statistical Methods and Applications 17 (1): 113–41. https://doi.org/10.1007/s10260-007-0061-9.
41. Rahman, Mizanur, Deluar J. Moloy, and Sifat Ar Salan. 2021. “Evaluating the Approaches of Small Area Estimation Using Poverty Mapping Data.” Journal of Statistical and Econometric Methods 10 (2): 1–21. https://doi.org/10.47260/jsem/1021.
42. Rao, JNK, and Isabel Molina. 2015. *Small Area Estimation*. 2nd ed. Hoboken.
43. Shi, Chengchun. 2018. Package ‘ BayesSAE,’ issued 2018.
44. Tadesse, Sebsibe, and Yinges Alemu. 2015. “Urban-Rural Differentials in Child Undernutrition in Ethiopia.” International Journal of Nutrition and Metabolism 7 (1): 15–23. https://doi.org/10.5897/IJNAM2014.0171.
45. Tekile, Abay Kassa, Ashenafi Abate Woya, and Garoma Wakjira Basha. 2019. “Prevalence of Malnutrition and Associated Factors among Under-Five Children in Ethiopia: Evidence from the 2016 Ethiopia Demographic and Health Survey.” BMC Research Notes 12 (1): 1–6. https://doi.org/10.1186/s13104-019-4444-4.
46. Teshome, B, W Kogi-Makau, Z Getahun, and G Taye. 2010. “Magnitude and Determinants of Stunting in Children under Five Years of Age in Food Surplus Region of Ethiopia: The Case of West Gojam Zone.” Ethiopian Journal of Health Development 23 (2): 98–106. https://doi.org/10.4314/ejhd.v23i2.53223.
47. Trevisani, Matilde, and Nicola Torelli. 2007. “Hierarchical Bayesian Models for Small Area Estimation with Count Data.” *Dipartimento Di Scienze Economiche e Statistiche, Universita Degli Studi Di Trieste*.
48. ———. 2017. “A Comparison of Hierarchical Bayesian Models for Small Area Estimation of Counts.” *Open Journal of Statistics* 07 (03): 521–50. https://doi.org/10.4236/ojs.2017.73036.
49. UNICEF, WHO, and World Bank. 2020. “Levels and Trends in Child Malnutrition: Key Findings of the 2020 Edition of the Joint Child Malnutrition Estimates.” Geneva: WHO 24 (2): 1–16.
50. Woldie, Mirkuzie, Challi Jirra, and Girma Azene. 2011. “Presence and Use of Legislative Guidelines for the Distribution of Decentralized Decision Making Authority in the Jimma Zone Health System, Southwest Ethiopia.” Ethiopian Journal of Health Sciences 21 (Suppl 1): 29.

51. Woodruff, Bradley A, James P Wirth, Adam Bailes, Joan Matji, Arnold Timmer, and Fabian Rohner. 2017. “Determinants of Stunting Reduction in Ethiopia 2000–2011.” Maternal & Child Nutrition 13. https://doi.org/10.1111/mcn.12307.
52. Yeshaw, Yigizie, Sewnet Adem Kebede, Alemneh Mekuriaw Liyew, Getayeneh Antehunegn Tesema, Chilot Desta Agegnehu, Achamyeleh Birhanu Teshale, and Adugnaw Zeleke Alem. 2020. “Determinants of Overweight/Obesity among Reproductive Age Group Women in Ethiopia: Multilevel Analysis of Ethiopian Demographic and Health Survey.” BMJ Open 10 (3): e034963. https://doi.org/10.1136/bmjopen-2019-034963.
53. Yilema, S A, Y A Shiferaw, T Zewotir, and E K Muluneh. 2022a. “Improving Survey Based Estimates of Malnutrition Using Small Area Estimation.” Statistical Journal of the IAOS 38: 1–11. https://doi.org/10.3233/SJI-210892.
54. ———. 2022b. “Multivariate Small Area Estimation of Undernutrition for Children under Five Using Official Statistics.” *Statistical Journal of the IAOS* 38 (2): 1–12. https://doi.org/10.3233/SJI-220935.
55. ———. 2022c. “Spatial Small Area Estimates of Undernutrition for under Five Children in Ethiopia via Combining Survey and Census Data.” *Spatial and Spatio-Temporal Epidemiology* 42. https://doi.org/10.1016/j.sste.2022.100509.
56. You, Yong, and Beatrice Chapman. 2006. “Small Area Estimation Using Area Level Models and Estimated Sampling Variances.” Survey Methodology 32 (1).
57. You, Yong, and Qian M. Zhou. 2011. “Hierarchical Bayes Small Area Estimation under a Spatial Model with Application to Health Survey Data.” Survey Methodology 37 (1): 25–37.
58. Zhang, Junni L., and John Bryant. 2020. “Fully Bayesian Benchmarking of Small Area Estimation Models.” Journal of Official Statistics 36 (1): 197–223. https://doi.org/10.2478/jos-2020-0010.

Figures

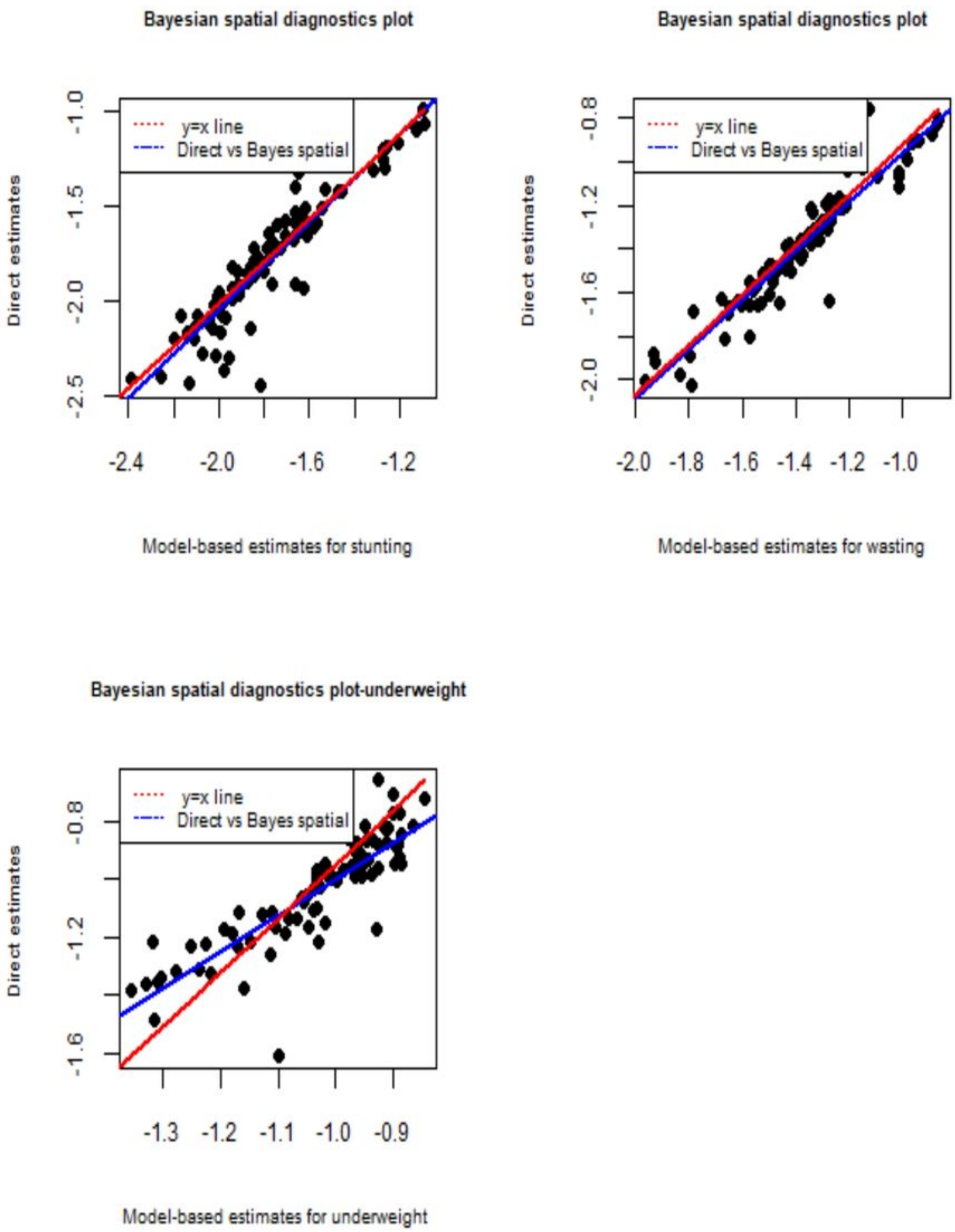

Figure 1

Bias diagnostic plots of stunting, wasting and underweight: Bayesian spatial SAE with *y = x* line (red line).

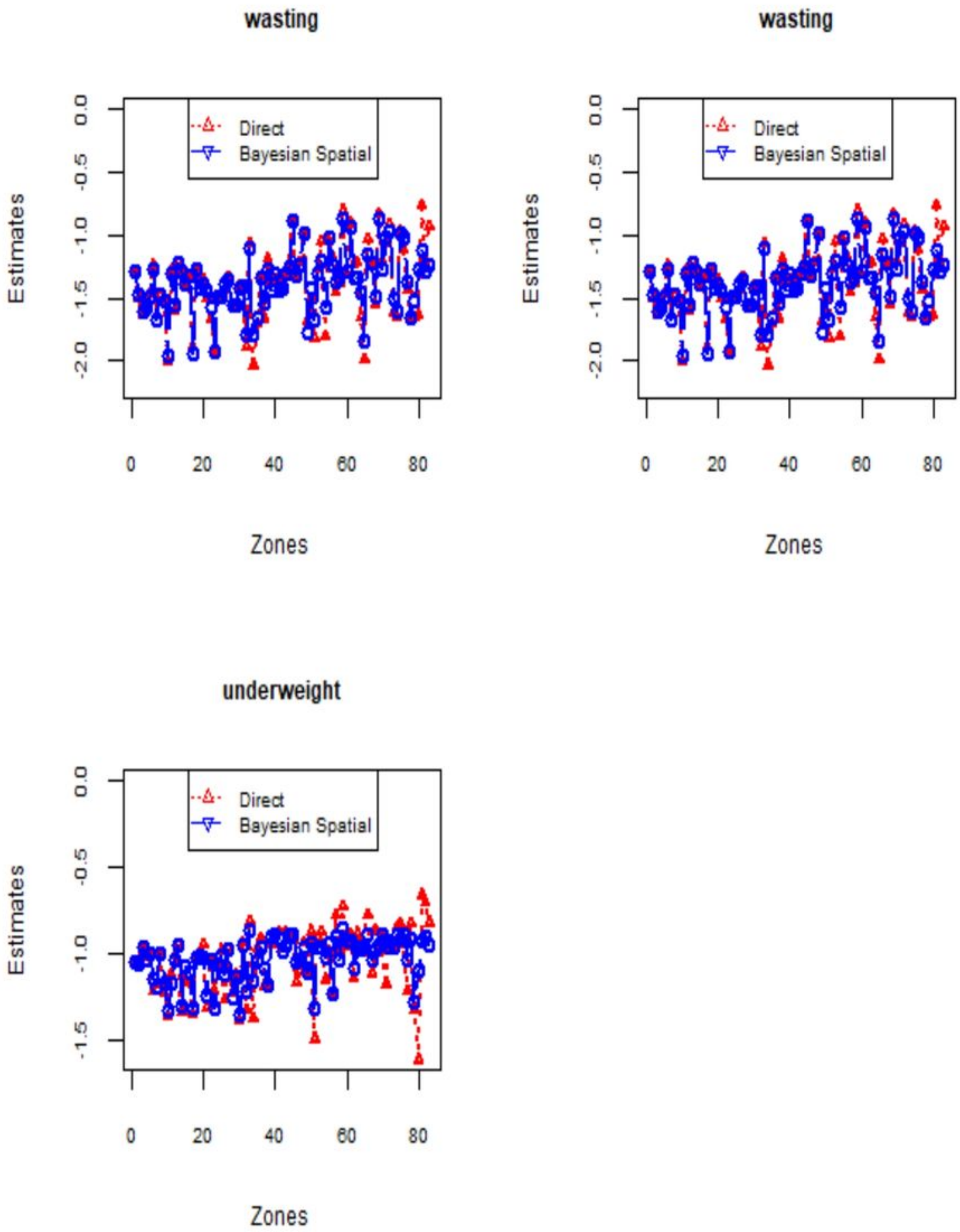

Figure 2

Zonal-wise direct and Bayesian spatial small area estimates

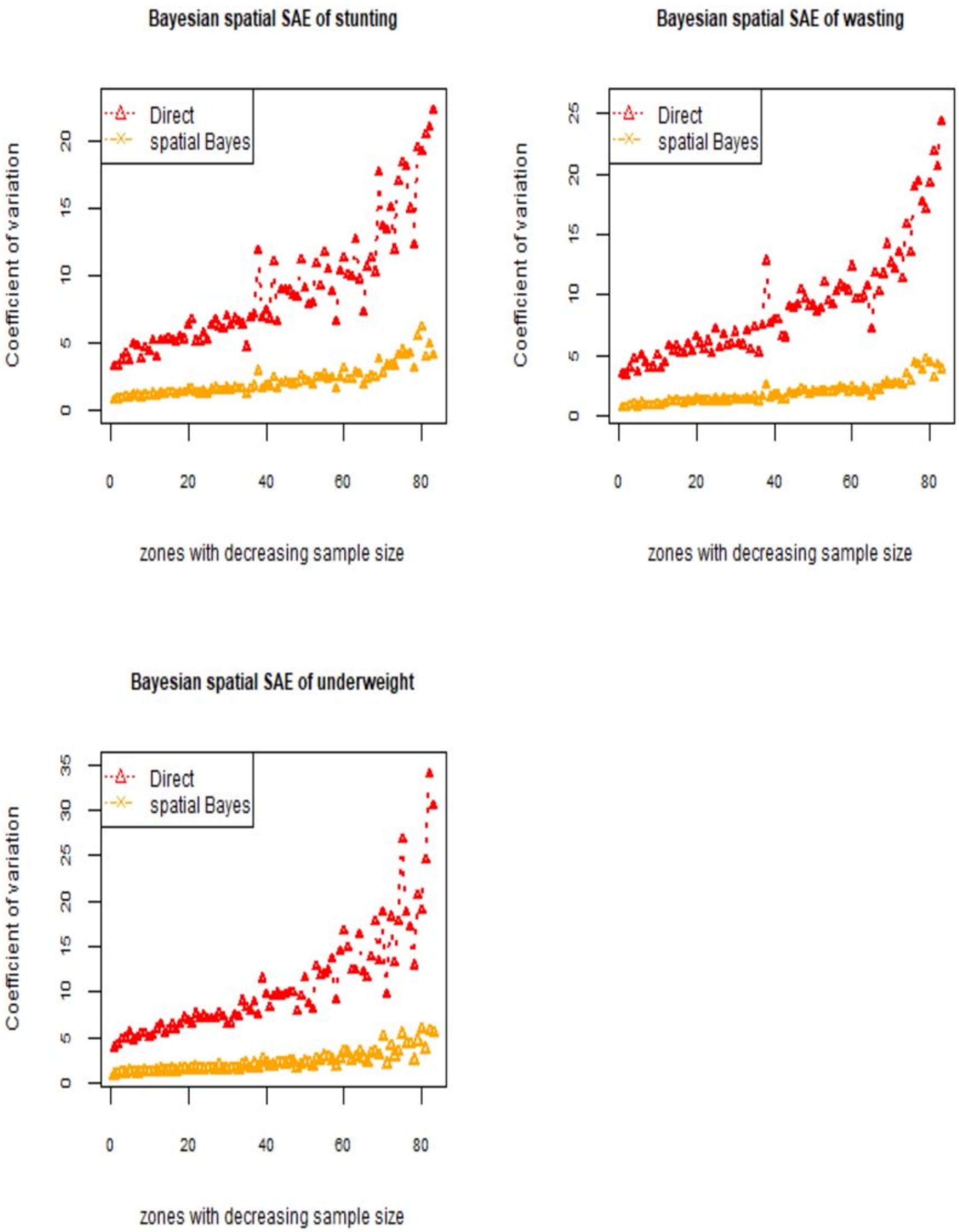

Figure 3

percentage CV of direct and model-based estimates for stunting, wasting and underweight

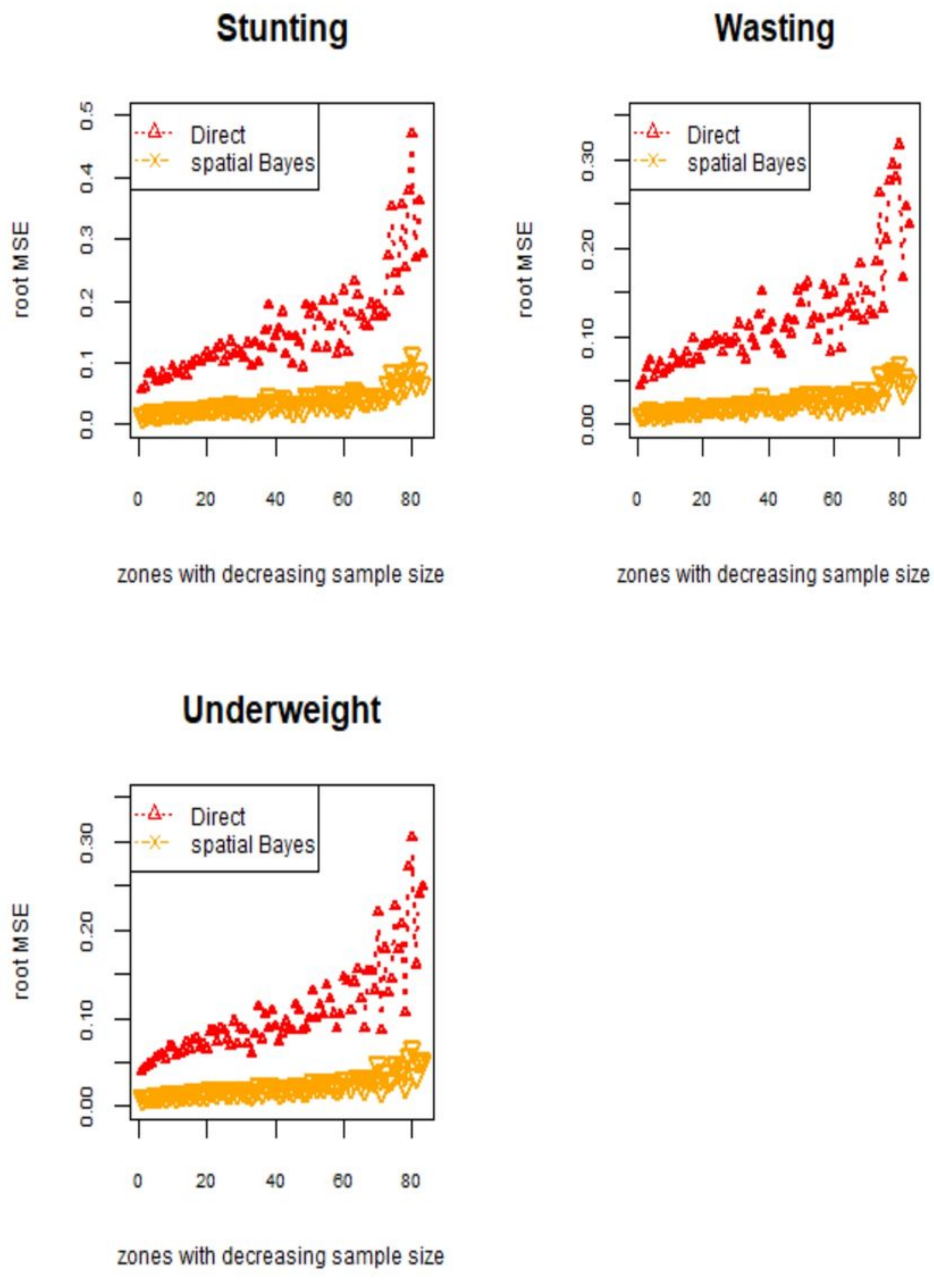

Figure 4

The root MSE of undernutrition with decreasing sample sizes

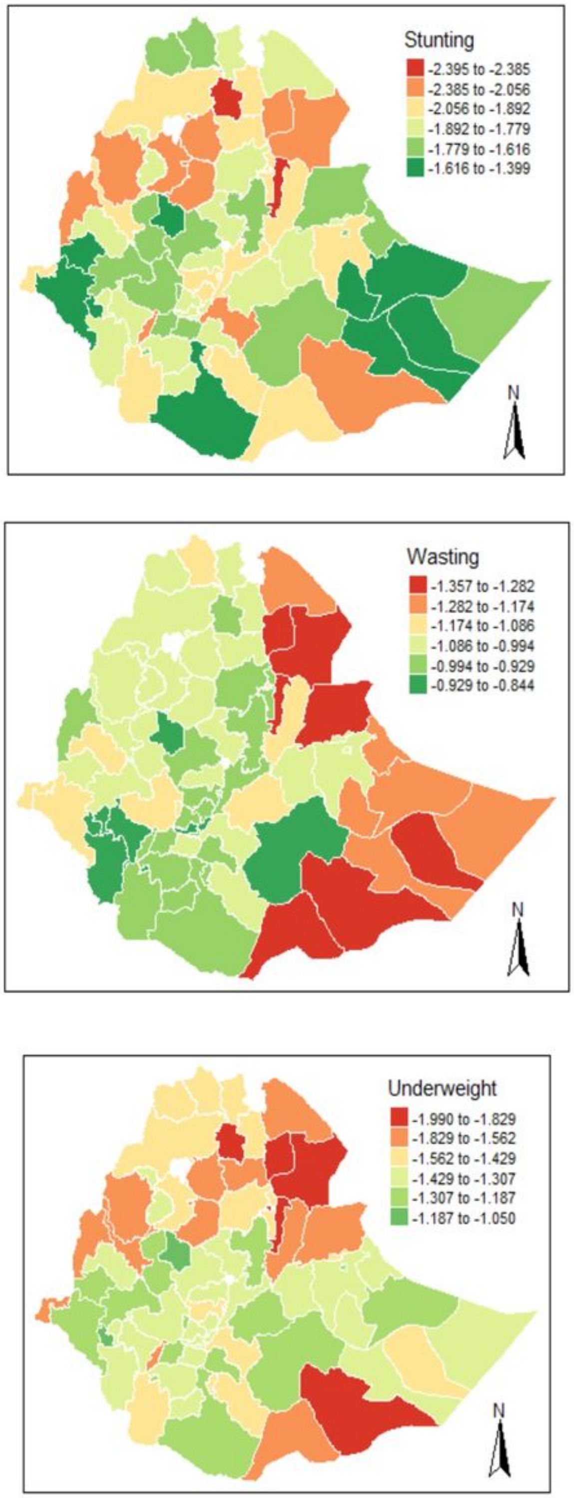

Figure 5

Model-based Bayesian spatial Fay-Herriot model undernutrition estimates for Ethiopian zones.