InPrime: Indonesian Journal of Pure and Applied Mathematics
Vol. 7, No. 2 (2025), pp. 128 – 143, doi: 10.15408/inprime.v7i2.46491
p-ISSN 2686-5335, e-ISSN: 2716-2478

Spatial Bayesian Small Area Estimation of Stunting Prevalence at the Subdistrict Level

Ita Wulandari[1]*, Ede Surya Darmawan[2] and Sofi Zamzanah[3]
[1,3]Departmen of Applied Statistics, Politeknik Statistika STIS, Jakarta, Indonesia
[2]Department of Health Administration and Policy, Faculty of Public Health,
Universitas Indonesia, Depok, Indonesia
Email: *ita.wulandari@stis.ac.id

Abstract

Achieving Sustainable Development Goal (SDG) 3—ensuring healthy lives and promoting well-being for all—requires reliable health indicators at disaggregated levels. However, conventional large-scale surveys often fail to capture local disparities due to limited sample sizes, leading to unreliable small-area estimates. This study proposes a hierarchical Bayesian Small Area Estimation (HB SAE) framework using a beta distribution and spatial Conditional Autoregressive (CAR) priors to estimate sub-district-level stunting prevalence in Banten Province, Indonesia. The model accounts for both the bound nature of prevalence data and spatial dependence across neighboring areas. Using data from the 2023 Indonesian Health Survey and auxiliary variables from the 2024 Village Potential Data (Podes), the proposed approach demonstrates improved analytical performance. Compared to direct estimation, the HB SAE model reduced the number of sub-districts with Relative Standard Errors (RSE) above 25% from 55 to 44, achieving lower Mean Squared Error (MSE) values and more stable estimates. These results confirm that incorporating spatial effects enhances the precision and reliability of stunting prevalence estimates. The findings provide analytically grounded evidence for localized, data-driven interventions to accelerate efforts to reduce stunting and strengthen policy planning at the sub-district level.

Keywords: Bayesian hierarchical model; small area estimation; spatial analysis; stunting prevalence; CAR model.

Abstrak

Pencapaian Sustainable Development Goal *(SDG) 3, yaitu menjamin kehidupan yang sehat dan meningkatkan kesejahteraan bagi semua orang, memerlukan indikator kesehatan yang andal hingga tingkat wilayah kecil. Namun, survei berskala besar sering kali tidak mampu menggambarkan variasi lokal akibat keterbatasan ukuran sampel, sehingga menghasilkan estimasi yang tidak stabil pada wilayah kecil. Penelitian ini mengusulkan kerangka kerja* Hierarchical Bayesian Small Area Estimation (HB SAE) *menggunakan distribusi beta dengan* prior spasial Conditional Autoregressive (CAR) *untuk mengestimasi prevalensi stunting di tingkat kecamatan di Provinsi Banten, Indonesia. Model ini mempertimbangkan sifat terbatas (bounded) dari data prevalensi dan ketergantungan spasial antarwilayah. Dengan memanfaatkan data Survei Kesehatan Indonesia 2023 serta variabel bantu dari data Potensi Desa (Podes) 2024, model HB SAE menunjukkan kinerja analitis yang lebih baik. Dibandingkan dengan estimasi langsung, model HB SAE menurunkan jumlah kecamatan dengan Relative Standard Error (RSE) di atas 25% dari 55 menjadi 44, dengan nilai Mean Squared Error (MSE) yang lebih rendah dan hasil estimasi yang lebih stabil. Hasil ini menunjukkan bahwa penambahan efek spasial meningkatkan ketepatan dan reliabilitas estimasi prevalensi stunting. Temuan ini memberikan dasar analitis yang kuat untuk mendukung intervensi berbasis data dalam percepatan penurunan stunting dan perencanaan kebijakan di tingkat kecamatan.*

Kata Kunci: *Model hierarkis Bayes; Small area estimation; Analisis spasial; Prevalensi stunting; Model CAR.*

2020MSC: 62F15, 62M30.

*) Corresponding author
Submitted September 17th, 2025, Revised October14th, 2025,
Accepted for publication October 15th, 2025, Published Online October 17th, 2025

1. INTRODUCTION

Reliable health statistics at disaggregated levels are essential for capturing disparities that are often obscured within national or provincial aggregates. This need is particularly evident in the context of the third Sustainable Development Goals (SDGs), which aim to ensure healthy lives and promote well-being for all people by 2030 [1]. Among the health challenges that threaten progress toward this goal, child stunting remains a persistent concern in many low- and middle-income countries. Stunting, a condition caused by chronic undernutrition, not only increases the risk of morbidity and mortality but also impairs cognitive development, reduces productivity, and undermines human capital formation, thereby perpetuating intergenerational cycles of poverty and inequality [2].

In Indonesia, stunting continues to represent a critical public health issue. While large-scale surveys, such as the Indonesian Health Survey and the Demographic and Health Survey (DHS), provide robust estimates at the provincial and national levels, they are less effective in identifying disparities within smaller administrative areas. Disaggregated information, however, is indispensable for designing targeted and context-specific interventions at lower levels such as districts, sub-districts, and even villages [3]. Traditional direct estimation methods are inadequate for this purpose because of small sample sizes in many domains, which result in unreliable estimates and wide margins of error [4], [5].

To overcome these limitations, Small Area Estimation (SAE) has emerged as a robust statistical framework for addressing these challenges. SAE enables the generation of precise estimates for small areas by integrating survey data with auxiliary sources of information, thereby improving precision without increasing survey costs [6]. Recent methodological developments highlight the advantages of hierarchical Bayesian (HB) approaches to SAE, which can incorporate prior knowledge, accommodate complex data structures, and provide full posterior distributions that allow for comprehensive quantification of uncertainty [7], [8], [9]. In this study, the methodological contribution lies in integrating the beta distribution, which is particularly suitable for modeling bounded proportions such as stunting prevalence, with spatial Conditional Autoregressive (CAR) priors that capture geographic dependencies across sub-districts. This combination extends beyond a direct application of existing models, representing a modification that enhances SAE's ability to address both the distributional characteristics of the response variable and spatial correlation among areas. Such an approach improves the accuracy and stability of the resulting estimates compared to standard SAE models without spatial or distributional adjustments [10], [11].

Empirical studies support the utility of spatial Bayesian SAE in public health applications. For instance, research in Ethiopia demonstrated that spatial Bayesian SAE improved the reliability of estimates for nutritional indicators such as stunting, wasting, and underweight, producing mean squared errors substantially lower than those obtained through direct estimation methods [12], [13], [14]. In Indonesia, spatial and spatio-temporal modeling approaches have also been applied to identify determinants of stunting and to map high-risk areas, providing critical evidence for localized interventions [3].

The urgency of generating reliable health statistics is further underscored by Indonesia's Vision 2045, which aspires to position the country among the world's developed nations by its centennial of independence. A cornerstone of this vision is the development of high-quality human capital, and reducing stunting is recognized as a strategic priority since a stunting-free generation will be healthier, more productive, and globally competitive [15], [16]. Producing precise child health statistics at the district and sub-district levels is therefore critical to supporting evidence-based policymaking and accelerating progress toward both SDG 3 and the broader agenda of Golden Indonesia 2045.

This study applies a hierarchical Bayesian SAE model with spatial effects to estimate the prevalence of stunting at the district level in Banten Province, Indonesia. The objectives are threefold: (i) to generate small-area prevalence estimates that are more reliable than direct survey estimators, (ii) to capture geographic heterogeneity and identify high-prevalence areas through spatial modeling, and (iii) to provide actionable evidence to guide policymakers in designing targeted interventions. In doing so, this study contributes methodologically by advancing the application of spatial Bayesian SAE in health statistics and practically by informing Indonesia's efforts to reduce stunting as part of its SDG commitments and its long-term aspiration of achieving a Golden Indonesia by 2045.

2. METHODS

2.1. Direct Estimation

Direct estimation is a method of estimating the value of parameters such as averages, proportions, or totals for a region or domain based on sample data obtained directly from that area. The determination of stunting status in children aged 0–59 months is carried out through the calculation of $z - scores$, which are obtained by considering the factors of gender, age, and height of the child. The median height of the reference standard and the standard deviation value of the reference were obtained from the child anthropometric standards in Indonesia which refer to the WHO Child Growth Standards [17].

$$Z - Score = \frac{Individual\ height - Reference\ standard\ median\ height}{Reference\ standard\ deviation\ value}.$$

The z-score value is then the basis for the classification of nutritional status according to the provisions of the Minister of Health of the Republic of Indonesia Number 2 of 2020 [18].

Table 1. Categories and nutritional status thresholds according to H/A (Height for Age)

Nutrition Status Categories	Thresholds ($Z - Score$)
Very short (severely stunted)	< -3 SD
Short (stunted)	-3 SD to -2 SD
Normal	-2 SD to +1SD
Tall	>+1 SD

Source: Regulation of the Minister of Health of the Republic of Indonesia No. 2 of 2020

The calculation of the prevalence of stunting can directly use the following formula:

$$\hat{\theta}_i = \frac{\sum_{i=1}^{n\cup} y_{ij} \times w_{ij}}{\sum_{i=1}^{n} w_{ij}} \times 100\%; \qquad i = 1, 2, \dots, m, \tag{1}$$

where $\hat{\theta}_i$ is an estimate of the prevalence of stunting in a sub-district i, y_{ij} is the stunting status of the j child in a sub-district i, and w_{ij} is the individual weigher of the j child in a sub-district i. If the child is stunted, then $y_{ij} = 1$, while if the child is not stunted, then $y_{ij} = 0$.

This research aims to estimate the prevalence of stunting at the sub-district level in Banten Province. Banten Province consists of 4 districts and 4 cities. At the sub-district level, Banten Province has 155 sub-districts. This study will estimate the prevalence of stunting in the 155 sub-districts. The results of the direct estimation of stunting prevalence were obtained from the 2023 Indonesian Health Survey (SKI) obtained from the Indonesian Ministry of Health.

2.2. Indirect Estimation

2.2.1. Data Exploration

The data obtained from direct estimation is then explored to understand its characteristics thoroughly. The exploration process includes checking the range of values to find out the variation of the data, distribution tests to identify whether the data is distributed normally or not, and analysis of the existence of missing data that can affect the accuracy of the estimation results. In addition, the Moran's I test was also carried out to detect the presence of spatial autocorrelation between observation units. The findings from this exploration stage are an important basis for determining the indirect estimation model that best suits the data structure.

2.2.2. Variable Selection

In addition to using data from SKI, this study also uses the 2024 Village Potential Data (Podes) obtained from the Indonesian Central Statistics Agency (BPS). Stunting prevalence data from SKI was used as a response variable, while Podes data was used as a predictor variable. The details of the variables used in this research are described in Table 2 [19], [20]. All prospective predictor variables were determined based on previous research and BPS publications related to stunting in Indonesia.

Table 2. Research variables

Dimension	Variable Code	Variable Name
-	Y	Direct estimation of stunting prevalence
Health	X_1	Ratio of referral health facilities (advanced) per 1000 population
Health	X_2	Ratio of first-level (primary) health facilities per 1000 population
Health	X_3	Ratio of maternity health facilities per 1000 population
Health	X_4	Ratio of pharmaceutical facilities per 1000 population
Health	X_5	Ratio of active posyandu per 1000 population
Health	X_6	Ratio of health workers living in villages per 1000 inhabitants
Stunting services	X_7	Ratio of supplementary feeding services (PMT) Counseling
Stunting services	X_8	PMT ratio of pregnant women with chronic energy deficiency or high risk from poor families
Stunting services	X_9	Ratio of training activities for pregnant women
Stunting services	X_{10}	Ratio of training activities for mothers of toddlers
Residential environment	X_{11}	Ratio of safe drinking water access services
Residential environment	X_{12}	Ratio of healthy toilet access services
Residential environment	X_{13}	Ratio of slum families per 1000 population
Residential environment	X_{14}	Ratio of mobile phone tower facilities or Base Transceiver Station (BTS)
Food	X_{15}	Ratio of food insecurity sufferers per 1000 population
Food	X_{16}	Ratio of the number of malnourished residents of marasmus and kwashiorkor per 1000 population
Education	X_{17}	Ratio of Early Childhood Education (PAUD) per 1000 population
Education	X_{18}	Ratio of health insurance services for pregnant women from poor families
Education	X_{19}	Ratio of health insurance services for clown children from poor families

Variable selection was carried out in two stages, namely the stepwise regression method on all predictor variable candidates and backward elimination which is based on the credible interval value. Stepwise regression is a forward selection process that checks back with a backward elimination process at each step across all independent variables entered into the model [21]. Then, the results of the selected variables are checked for significance based on the credible interval value. Variables are said to be significant at a significance level of 5 percent when the credible interval at intervals of 2,5 percent to 97,5 percent does not pass zero.

2.2.3. Hierarchical Bayes Small Area Estimation (HB SAE)

Small Area Estimation (SAE) is a method to estimate population parameters in a small area in conditions of insufficient direct data [6]. There are two types of models in SAE, namely area level model and unit level model, depending on the availability of accompanying variables. The basic area level model is written as follows.

$$\hat{\theta}_i = \boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta} + b_i v_i + e_i\,; \qquad i = 1, 2, \dots, m, \tag{2}$$

where $\hat{\theta}_i$ is the parameter estimation symbol for area i, $\boldsymbol{x}$ is a predictor variable, $\boldsymbol{\beta}$ is the vector of the model coefficient, $b_i v_i$ represents a random effect for area i, and e_i indicates measurement errors or noise that is not explained by the model. Some of the approaches that can be used in the area level SAE method are Best Linear Unbiased Prediction (BLUP), Empirical Best Linear Unbiased Prediction (EBLUP), Empirical Bayes (EB), and Hierarchical Bayes (HB) [6], [22]. The HB method was chosen because it takes into account diversity in hyperparameters, has more direct model specifications for various sources of variation, and offers clearer inferences with easier computation through MCMC techniques [23].

The Bayesian method is a method of parameter estimation that combines knowledge related to the distribution of unknown parameters (prior) with information from sample data (likelihood) into posterior distributions [24]. The posterior distribution serves as the main reference in the inference process in the Bayesian method. To overcome the complexity of integration, the Markov Chain Monte Carlo (MCMC) algorithm can be applied, which is a Bayesian computational method that generates random variables through the Markov chain [25], [26]. In this study, the HB SAE process was carried out using the hbsaems package in the R application. This package implements Stan in doing HB SAE. Stan is a programming language and statistical platform for Bayesian inference that uses the No-U-Turn Sampler (NUTS), an adaptive variant of the HMC [27]. Stan was developed to overcome the limitations of computational and algorithmic efficiency in BUGS and JAGS [27], [28], [29].

One of the distributions used in modeling data that has a value range between 0–1 is the beta distribution [30]. The HB beta model in SAE uses the beta distribution in the sampling model and the logit link function in the linking model [6], [30].

i. Sampling model

$$\hat{\theta}_i | \theta_i \sim Beta(a_i, b_i), \qquad i = 1, \dots, m, \tag{3}$$

with $a_i = \theta_i \left(\frac{n_i}{deff_{iw}} - 1\right)$ and $b_i = (1 - \theta_i)\left(\frac{n_i}{deff_{iw}} - 1\right)$. θ_i is the parameter of probability, n_i is the number of samples i, and $deff_{iw}$ is the effect of sampling design.

ii. Linking model

$$logit(\theta_i) | \boldsymbol{\beta}, \sigma_v^2 \overset{iid}{\sim} N\left(\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}, \sigma_v^2\right), \qquad i = 1, \dots, m. \tag{4}$$

The model states that the logical transformation of the probability θ_i, follows a normal distribution with an average of $\boldsymbol{x}_i^T\boldsymbol{\beta}$ and a variance of σ_v^2. $\boldsymbol{\beta}$ is the regression coefficient vector, $\boldsymbol{x}_i$ is the covariate vector for the unit-i, and σ_v^2 indicates a random variation of components.

In doing HB SAE, it can be done by paying attention to the spatial effects that sometimes appear on the data. One method that can be applied is the Conditional Autoregressive Model (CAR). Inter-region dependencies are modeled by entering spatial information through a neighborhood matrix (adjacency matrix), so that the estimation of an area is influenced by the characteristics of the surrounding area. This model can then be developed into:

$$\theta_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + v_i + u_i, \tag{5}$$

where u_i expresses the addition of the spatial random effect of the CAR model obtained through the following formula.

$$u_i|u_{-i} \sim N\left(\frac{\sum_{j\neq i} w_{ij}u_j}{\sum_{j\neq i} w_{ij}}, \frac{\sigma_u^2}{\sum_{j\neq i} w_{ij}}\right). \tag{6}$$

Formula 8 shows that the u_i value depends on the u_j value of the neighboring area j. The spatial weight of w_{ij} represents the proximity or connection between areas i and j. Thus, u_i captures spatial effects, so adjacent areas tend to have similar characteristics.

2.2.4. Convergence Checking

Convergent conditions occur when the MCMC algorithm has reached its equilibrium state (stationary state). The convergence in the MCMC algorithm can be seen through a visual approach with a diagnostic plot consisting of [31]:

1. Trace Plot, which is a graph of the values of the generated parameters (y-axis) and the iteration values used (x-axis). The equilibrium condition of MCMC occurs when the shape of the trace plot does not have a periodic pattern because the average and variety are relatively constant.
2. Density Plot, which is the distribution of the values of the parameters that have been generated. Density plots that have converged generally look like a bell curve.
3. Autocorrelation Plot, which is a graph that shows the correlation between samples for each iteration shown through the graph the autocorrelation function (ACF). Visually, MCMC is said to be convergent when the ACF graph is cut-off.

2.3. Mapping Estimated Results

The results of the estimated stunting prevalence in Banten Province were then mapped. This is done to provide a visual overview of the distribution and variation in stunting prevalence between regions. With mapping, stunting rate data in the form of statistical numbers can be translated into spatial information that is easier to understand and analyze.

3. RESULTS

3.1. Data Exploration and Variable Selection

The sub-district level prevalence of stunting in Banten Province ranges between 0 and 1, with no districts exhibiting exact values of 0 or 1. The data exhibit non-normal distribution, and two districts lack direct estimation results due to missing data. Furthermore, the Moran's I test reveals significant spatial autocorrelation (p-value < 0.05), justifying the incorporation of spatial effects in the estimation process. An examination of the Relative Standard Errors (RSE) of direct estimates indicates that 55 sub-districts have RSE values exceeding 25%, suggesting low reliability and high uncertainty in these estimates. Based on these exploratory findings, the Hierarchical Bayes Small Area Estimation approach with beta distribution and spatial effects is deemed suitable for estimating stunting prevalence at the district level in Banten Province. A statistical summary of the direct estimates can be seen in Table 3.

Table 3. Summary of statistics for direct estimation

Statistics	Direct Estimate (Percent)	RSE (Percent)
Min	3,06	0,00
Q1	16,09	10,36
Median	24,17	20,31
Mean	26,55	24,45
Q3	34,50	31,74
Max	71,36	103,94
NA	2	2

The distribution of stunting prevalence across sub-districts in Banten Province displays marked heterogeneity, underscoring pronounced disparities in nutritional outcomes within the region. Notably, the interquartile range reveals that a substantial proportion of sub-districts exhibit prevalence levels exceeding the provincial average, highlighting persistent and entrenched inequalities in child health outcomes. Figure 1 shows the distribution of RSE values for direct estimation of stunting prevalence at the sub-district level. The visualization is divided into three categories based on the RSE value from the estimated results, namely below 25 percent, above 25 percent, and NA.

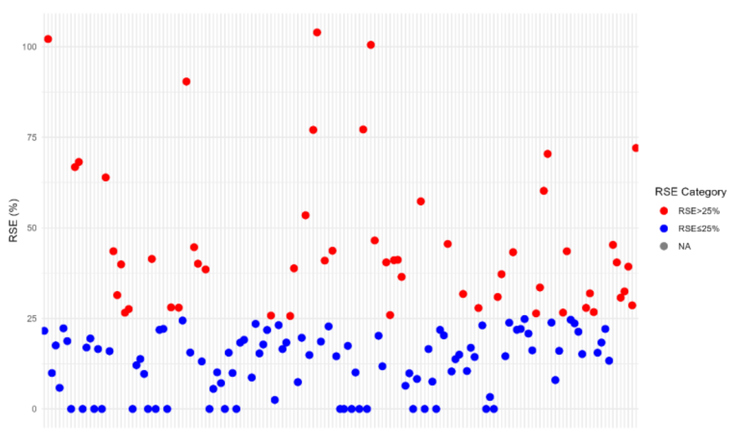

Figure 1. Distribution of RSE values for direct estimation of stunting prevalence

Two sub-districts (Lebak Gedong and Koroncong) were not estimable due to insufficient sample data. Notably, from a precision standpoint, more than a third of the sub-districts exhibited Relative Standard Errors (RSE) exceeding the 25% threshold established by Statistics Indonesia, with 55 sub-districts specifically showing RSE values above this threshold (see Figure 1). This result highlights the limited reliability of direct estimates for numerous sub-districts, underscoring the imperative of adopting model-based approaches to generate more stable and policy-relevant small-area statistics. We elaborate on this issue further in the next subsection.

Predictor selection was performed using a hybrid approach combining stepwise regression and backward elimination to identify the most parsimonious model with optimal predictive performance. The stepwise procedure iteratively introduced variables, while backward elimination refined the model by removing non-significant predictors that fell outside the 95% Bayesian credible interval (2.5%–97.5%). The process of each step in stepwise regression that explains what variables will enter and exit is shown in Table 4. The stepwise regression process stops at step 4 and gives the final result of the variables that affect stunting are $\mathbf{X_{17}}$, $\mathbf{X_5}$, $\mathbf{X_{11}}$, dan $\mathbf{X_2}$.

Table 4. Stepwise regression process

Process	Variables
1	X_{17}
2	$X_{17} + X_5$
3	$X_{17} + X_5 + X_{11}$
4	$X_{17} + X_5 + X_{11} + X_2$

The results of the variables from stepwise regression were then checked based on the credible interval. Further evaluation revealed that variable $\mathbf{X_2}$ had a 95% credible interval that encompassed zero, indicating a non-significant effect at the 5% level. Consequently, $\mathbf{X_2}$ was removed from the model, resulting in a final specification that retained three predictors: $\mathbf{X_{17}}$, $\mathbf{X_5}$, and $\mathbf{X_{11}}$. This parsimonious model ensures a balance between interpretability and statistical robustness, while mitigating the risk of overfitting.

3.2. Spatial Conditional Autoregressive Model

Sub-districts without estimated values were excluded from the analysis due to the absence of essential initial information required for model calibration and evaluation. Convergence of the MCMC process is critical to ensure the reliability of parameter estimates. To evaluate convergence, several diagnostic tools were employed, including trace plots, density plots, and autocorrelation plots. Based on the diaknostic results shown in Figure 2, it can be seen that the iteration has converged on the 15000th iteration. It is characterized by a Trace Plot that shows lines that overlap each other and move randomly around the center value, a Density Plot that shows a smooth, unimodal, and relatively symmetrical form of distribution, and an Autocorrelation Plot that shows a rapid decline towards zero at small lags that mean low autocorrelation.

Convergence was further confirmed by Rhat values for all parameters approaching 1, suggesting that the Markov chains successfully converged to the target distribution. In addition, the Effective Sample Size (ESS) values, including both Bulk_ESS and Tail_ESS, were relatively large for most parameters, indicating that a sufficient number of effective samples were obtained to ensure stable estimation and reliable inference. The large ESS values also imply that the MCMC simulation efficiently explored the parameter space, thereby reducing the risk of biased estimates. Overall, the HB SAE beta model with spatial effects demonstrated convergence after 15000 iterations, providing confidence in the reliability of parameter estimates and supporting the validity of subsequent inferences (see Table 5).

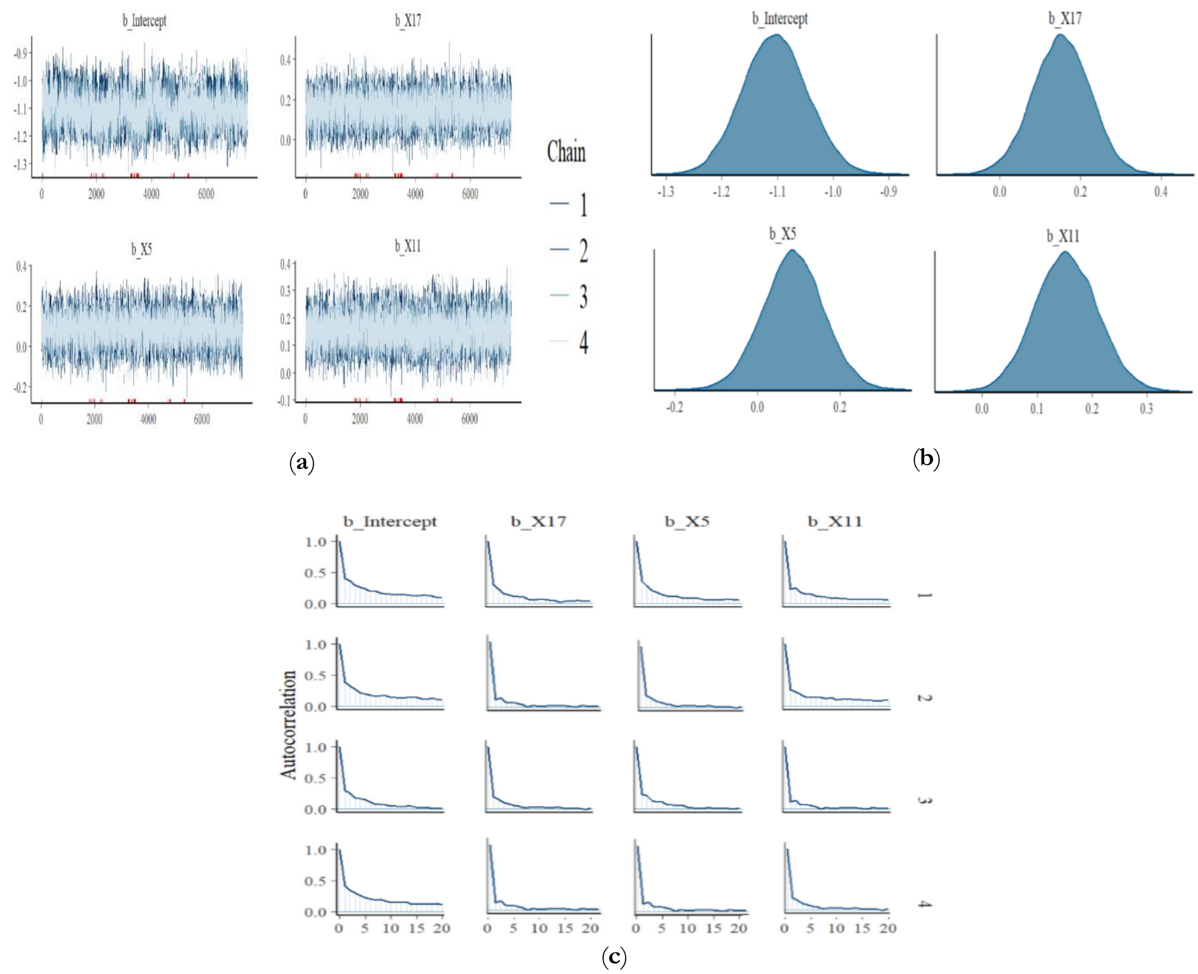

Figure 2. Diagnostics Visual MCMC in HB SAE beta distribution using spatial effects on stunting prevalence in Banten Province, (**a**) Trace Plot, (**b**) Density Plot, and (**c)** Autocorrelation Plot

Table 5. Rhat Value and Effective Sample Size in the spatial beta HB SAE model

	Rhat	**Bulk_ESS**	**Tail_ESS**
Intercept	1,00144	994,61	5102,53
X_{17}	1,00022	10546,64	15781,97
X_5	1,00082	6077,44	11131,27
X_{11}	1,00038	11107,93	17670,59

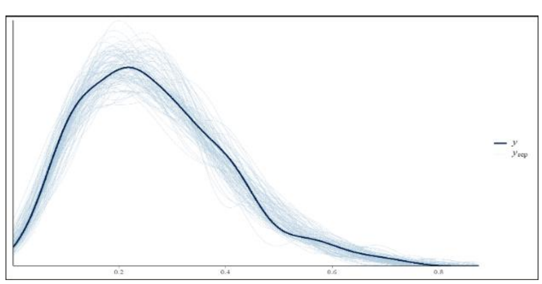

Figure 3. Posterior predictive checking

Posterior predictive checking was conducted to evaluate the adequacy of the model. As illustrated in Figure 3, the hierarchical Bayesian SAE beta model with spatial effects produces a posterior predictive distribution that closely aligns with the observed data, indicating that the model effectively captures the underlying structure and patterns. This alignment demonstrates that incorporating spatial effects improves the model's ability to represent local variations and dependencies, thereby enhancing the accuracy of small area estimates. Ensuring model adequacy is essential for reliable inference and decision-making, and the results confirm the suitability of the hierarchical Bayesian SAE beta model with spatial effects for this purpose. A statistical summary of the results of indirect estimation of stunting prevalence using HB SAE can be seen in Table 6.

Table 6. Comparison of the results of the estimated stunting prevalence

Statistics	Indirect Estimation (Percent)	RSE (Percent)
Min	8,52	12,96
Q1	18,91	17,89
Median	24,99	20,77
Mean	26,67	22,99
Q3	33,13	26,27
Max	59,61	54,34

Building on this validation, indirect estimates of stunting prevalence were generated at the sub-district level using the hierarchical Bayesian SAE beta model with spatial effects. The summary statistics in Table 6 indicate that the model-based approach produces estimates with markedly reduced variability and improved precision compared to direct estimation. In particular, the Relative Standard Errors (RSE) are substantially lower, with the majority of sub-districts falling below the 25% threshold commonly recommended for reliable publication. This gain in precision reflects the model's ability to borrow strength from neighboring sub-districts through spatial effects, thereby improving the reliability of estimates, especially in areas with small sample sizes. The observed reduction in RSE highlights the potential of the model-based approach to generate more accurate and policy-relevant insights, providing stronger evidence to guide interventions at the sub-district level.

3.3. Comparison of Results

The indirect estimation model using HB SAE was able to reduce the RSE compared with direct estimation. To assess the goodness of fit of the model, a comparison was conducted between the results of direct and indirect estimation. The comparison was made using the values of RSE and Mean Squared Error (MSE), since a direct comparison of the point estimates is not meaningful, as both are estimates rather than the actual population values. Table 7 shows the comparison of model evaluation results.

Based on Table 7, it is evident that estimation using the HB SAE beta model with spatial effects reduces both the RSE and MSE. Focusing on the RSE values, the HB SAE beta model with spatial effects decreases the number of districts with an RSE above 25 percent. Under direct estimation, 55 districts exhibit an RSE greater than 25 percent, whereas under the HB SAE beta model with spatial effects, this number decreases to 44 districts. This shows that indirect estimation in stunting prevalence is more suitable to be used and can overcome the weakness of direct estimation. Indirect estimation can improve the accuracy of the estimate.

Table 7. Comparison of model evaluation results

Statistik	RSE (Persen)		MSE (Persen)	
	Direct Estimatiom	Indirect Estimation	Direct Estimatiom	Indirect Estimation
Min	0,00	12,96	0,00000	0,00114
Q1	10,36	17,89	0,00082	0,00225
Median	20,31	20,77	0,00219	0,00278
Mean	24,45	22,99	0,00408	0,00336
Q3	31,74	26,27	0,00488	0,00368
Max	103,94	54,34	0,03134	0,01545
NA	2	-	2	-

Furthermore, in the case of direct estimation, there are two sub-districts with missing results for the estimates, RSE, and MSE. This occurs because no samples were collected in those sub-districts, making the estimation process infeasible. In other words, the observation units in these areas are not represented by the survey data, and as a result, the model cannot generate predictions or accuracy measures such as RSE and MSE. In contrast, indirect estimation using the HB SAE beta model with spatial effects is able to provide estimates for all districts. For districts not represented in the survey data, predictions are obtained by borrowing strength from information in neighboring areas.

Thus, the HB SAE beta with spatial effects was shown to be superior to direct estimates in terms of stability and completeness of results. The decrease in the number of sub-districts with high RSE values shows that this model is able to increase the reliability of estimates at the sub-district level. In addition, the model's ability to produce estimates in sub-districts that do not have a sample confirms that this method is able to overcome the limitations of limited or uneven survey data. To easily compare the results of RSE and MSE from direct and indirect estimates, a graph is made shown in Figure 4. It can be seen that the distribution of RSE and MSE results in both estimates shows that indirect estimates are superior to direct estimates.

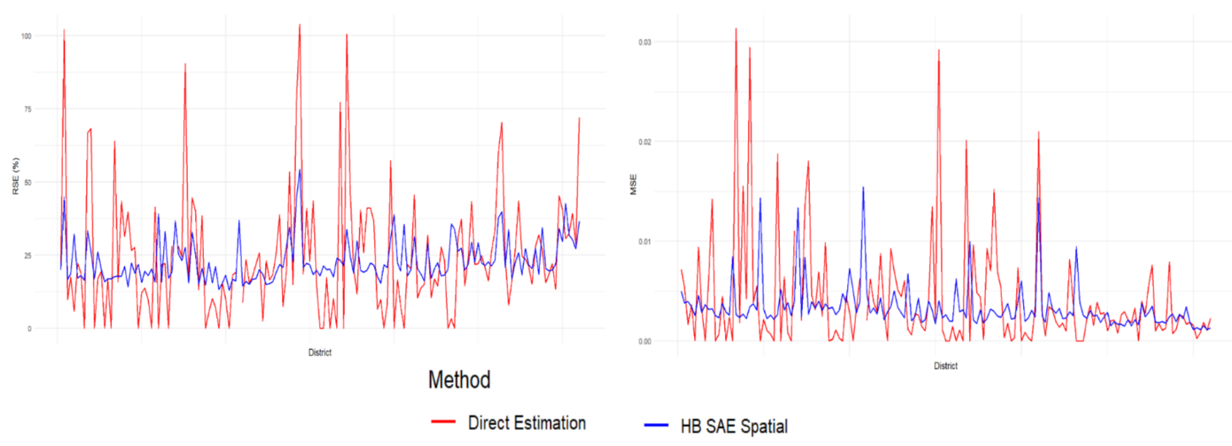

Figure 4. Visualization of model evaluation results: RSE (left) and MSE (right)

3.4. Geographical Patterns of Stunting

The estimation results for all sub-districts in Banten Province, obtained through direct estimation and indirect estimation using the HB SAE beta distribution with spatial effects, show that the indirect estimation produces relatively more stable and consistent results. Unlike direct estimation, which may generate NA values, the indirect estimation is able to provide estimates for all sub-districts. Furthermore, the estimation of stunting prevalence using the Spatial HB SAE method successfully reduces the number of sub-districts with high RSE values (above 25 percent). Mapping was carried out on the results of both estimates. Figure 5 is a visualization of the thematic map of the results of the estimation of the prevalence of stunting directly and indirectly in Banten Province which is grouped based on WHO standards.

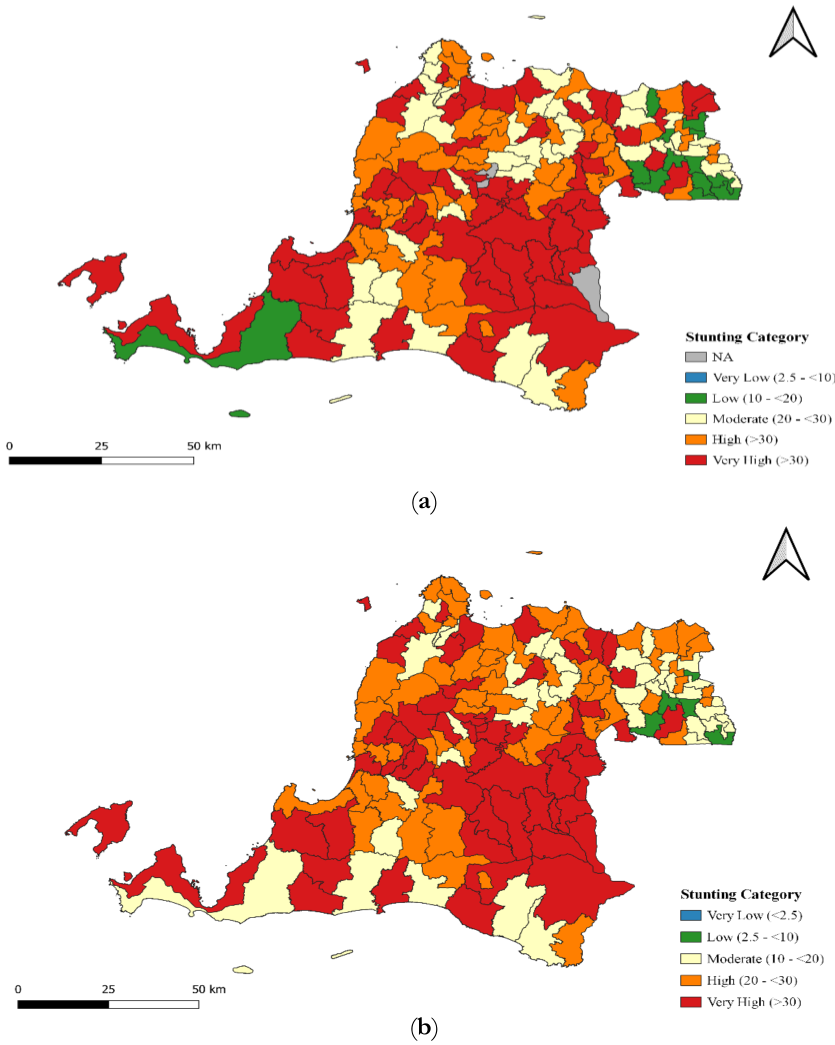

Figure 5. Mapping the results of the estimated prevalence of stunting in Banten Province, (**a**) Direct Estimation, (**b**) Indirect Estimation with HB SAE Spatial

Based on Figure 5, the results of direct estimation show a more diverse variety of categories, with areas that do not have an estimated value (NA) and the existence of several sub-districts with low prevalence. On the other hand, the results of indirect estimation through HB SAE Spatial produce a smoother and more consistent map, where the entire region can be estimated without any lost values, and the spatial pattern of stunting prevalence appears clearer with the dominance of the high to very high category.

Based on the results of the mapping, the majority of areas are red, it shows that stunting in Banten Province is in the very high category (stunting cases occur in more than 30 percent of the number of children in the region). This indicates that the problem of stunting is still a serious and widespread public health problem in Banten Province. This condition reflects significant challenges related to food security, parenting, maternal and child health, and uneven access to health services in many sub-districts. The dominance of the high to very high category shows that existing interventions are not sufficiently effective in reducing stunting rates, so a more targeted, region-based, and comprehensive policy strategy is needed to reduce the prevalence of stunting in a sustainable manner. Indirect estimation that increases the accuracy of estimates is expected to be the foundation for the government and the community in taking the right steps to deal with stunting in Banten Province.

4. DISCUSSION

This study demonstrates the effectiveness of the hierarchical Bayesian Small Area Estimation (HB SAE) with spatial effects in producing reliable sub-district level estimates of stunting prevalence in Banten Province. Compared with direct estimation, the model-based approach substantially reduced relative standard errors (RSE) and mean squared errors (MSE), thereby enhancing both the precision and stability of small-area statistics. These findings are in line with previous studies that highlight the superiority of model-based approaches in addressing challenges of limited sample sizes and high sampling variability [6], [8], [10].

The capacity of the spatial HB SAE model to generate estimates for sub-districts with no direct survey data underscores its practical relevance in contexts of incomplete or uneven data coverage. Similar evidence has been reported by Aswi et al. [3] in Indonesia and Srivastava et al. [4] in India, both of whom demonstrated that the incorporation of spatial modeling reduces estimation bias and improves reliability. Our findings further confirm that incorporating Conditional Autoregressive (CAR) priors is critical for capturing geographic patterns of stunting prevalence. Evidence of significant spatial autocorrelation suggests that child health outcomes are influenced by neighboring areas, consistent with the conclusions of You and Zhou [8] and more recent applications in Ethiopia and other low- and middle-income countries [12], [13].

Another important contribution of this study is the integration of the beta distribution within the HB SAE framework, which is particularly appropriate for bounded prevalence outcomes. This specification aligns with methodological advances that advocate distribution-specific approaches to improve model fit in SAE applications [7], [9], [11]. By borrowing strength across sub-districts, the model effectively mitigates unstable estimates in sparsely sampled areas and provides a more nuanced representation of regional disparities, reinforcing the findings of Gao and Wakefield [11] and Yilema et al. [13]. The use of Markov Chain Monte Carlo (MCMC) inference ensured convergence and robust parameter estimation, while posterior predictive checks confirmed the adequacy of the model.

Despite these contributions, several limitations should be acknowledged. The reliance on secondary survey data may introduce measurement error or reporting bias, and the CAR specification may not fully capture non-stationarity or more complex spatial dynamics [9]. Furthermore, the analysis is restricted to a cross-sectional design, and temporal variations in stunting prevalence were not considered. Future research should therefore extend this work by integrating longitudinal data, incorporating a richer set of socio-economic and environmental covariates, and exploring alternative spatial structures such as non-stationary or multivariate SAE models [9], [14]. These extensions would further strengthen methodological robustness and enhance the evidence base for localized policy interventions to reduce child stunting, thereby contributing to the achievement of Sustainable Development Goal 3.

5. CONCLUSIONS

This study confirms the effectiveness of the hierarchical Bayesian Small Area Estimation (HB SAE) model with spatial effects in producing reliable estimates of stunting prevalence at the sub-district level in Banten Province. Compared with direct estimation, the approach substantially reduces relative standard errors and mean squared errors, while overcoming data gaps in sub-districts without survey representation. The integration of the beta distribution enables an appropriate treatment of bounded prevalence outcomes, and the inclusion of Conditional Autoregressive (CAR) priors successfully captures geographic dependencies. The resulting estimates provide a more consistent and policy-relevant picture of stunting, highlighting marked inequalities across districts and offering valuable evidence for designing targeted interventions to accelerate progress toward Sustainable Development Goal 3 and Indonesia's Golden 2045 vision.

Reliance on secondary survey data may introduce measurement error or reporting bias, and the beta distribution imposes assumptions that may not fully capture the heterogeneity of stunting outcomes. Furthermore, the CAR prior assumes spatial stationarity, which may be restrictive in regions with more complex dynamics. Future research should therefore integrate explicit measurement error models, explore non-stationary and multivariate spatial frameworks, and extend the analysis to spatio-temporal settings. These methodological refinements would strengthen the robustness of small-area estimates, provide a richer understanding of child nutrition outcomes, and further enhance the utility of SAE methods for informing localized and evidence-based stunting reduction strategies.

REFERENCES

[1] Department of Economic and Social Affairs United Nations, "*Transforming our world: the 2030 Agenda for Sustainable Development*," https://sdgs.un.org/2030agenda (accessed October 1, 2025).

[2] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, and L. M. Neufeld, "A review of child stunting determinants in Indonesia," *Maternal and Child Nutrition*, vol. 14, no. 4. Blackwell Publishing Ltd, Oct. 2018. doi: 10.1111/mcn.12617.

[3] A. Aswi *et al.*, "Childhood stunting in Indonesia: assessing the performance of Bayesian spatial conditional autoregressive models," *Geospatial Health*, vol. 19, no. 2, July 2024, doi: 10.4081/gh.2024.1321.

[4] S. Srivastava, H. Chandra, S. K. Singh, and A. K. Upadhyay, "Mapping Changes in District Level Prevalence of Childhood Stunting in India 1998-2016: An Application of Small Area Estimation Techniques," *SSM - Popul. Health*, vol. 14, p. 100748, June 2021, doi: 10.1016/j.ssmph.2021.100748.

[5] A. Juliyanto, "Hierarchical Bayes Modeling in Small Area Estimation For Estimating Unemployment Proportion Under Complex Survey," PhD Thesis, Institut Teknologi Sepuluh Nopember, 2016.

[6] J. N. K. Rao and I. Molina, *Small area estimation*, Second. John Wiley & Sons, Inc., 2015.

[7] I. Wulandari, K. A. Notodiputro, A. Fitrianto, and A. Kurnia, "Hierarchical Bayesian Models for Small Area Estimation under Overdispersed Count Data," *Eng. Lett.*, vol. 31, no. 4, Dec. 2023.
[8] Y. You and Q. M. Zhou, "Hierarchical Bayes Small Area Estimation Under a Spatial Model with Application to Health Survey Data," *Surv. Methodol.*, vol. 37, pp. 25–37, June 2011.
[9] P. Anjoy and H. Chandra, "Hierarchical Bayes estimation of Small Area Means Under a Spatial Nonstationary Fay–Herriot Model," *Commun. Stat. - Simul. Comput.*, vol. 52, no. 7, pp. 3043–3061, July 2023, doi: 10.1080/03610918.2021.1926501.
[10] I. Molina and J. N. K. Rao, "Historical Overview of Small Area Estimation in the 50 th Birthday of the IASS," *Surv. Stat.*, vol. 88, pp. 23–35, 2023.
[11] P. A. Gao and J. Wakefield, "A spatial variance-smoothing area level model for small area estimation of demographic rates," Sept. 06, 2022, *arXiv*: arXiv:2209.02602. doi: 10.48550/arXiv.2209.02602.
[12] K. F. Muchie, A. K. Wanjoya, and S. M. Mwalili, "Small Area Estimation of Zone-Level Malnutrition among Children under Five in Ethiopia," *Math. Comput. Appl.*, vol. 27, no. 3, p. 44, May 2022, doi: 10.3390/mca27030044.
[13] S. A. Yilema, Y. A. Shiferaw, T. Zewotir, and E. K. Muluneh, "Multivariate Small Area Estimation of Undernutrition for Children Under Five Using Official Statistics," *Stat. J. IAOS*, vol. 38, no. 2, pp. 625–636, 2022, doi: 10.3233/SJI-220935.
[14] H. C. Chung and G. S. Datta, "Bayesian Spatial Models for Estimating Means of Sampled and Non-Sampled Small Areas," *Surv. Methodol.*, vol. 48, pp. 463–489, Dec. 2022.
[15] Ministry of National Development Planning of the Republic of Indonesia (Bappenas), *Roadmapof Sustainable Development Goals 2023-2030*. Deputy for Maritime Affairs and Natural Resources, Ministry of National Development Planning/ National Development Planning Agency, 2023.
[16] World Bank, *Investment Framework for Nutrition 2024: Overview*. World Bank Group, 2024.
[17] President of the Republic of Indonesia, "Peraturan Presiden Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting," 2021.
[18] Minister of Health of the Republic of Indonesia, "Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020." 2020.
[19] Central Statistics Agency (BPS), *Laporan Indeks Khusus Penanganan Stunting 2021-2022*. Jakarta: Badan Pusat Statistik, 2023.
[20] M. A. C. Hakim and S. Muchlisoh, "Penerapan Model Fay-Herriot Pada Estimasi Prevalensi Stunting Level Kecamatan di Nusa Tenggara Bara Tahun 2017," *Semin. NasionalOfficial Stat. 2019*, pp. 74–83, 2019.
[21] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, Eds., "Class Variables in Regression," in *Applied Regression Analysis*, in Springer Texts in Statistics. , New York: Springer-Verlag, 1998, pp. 269–323. doi: 10.1007/0-387-22753-9_9.
[22] R. A. Noviyanti, "Scan Statistic dengan Pendekatan Small Area Estimation Empirical Bayes untuk Mendeteksi Kantong Kemiskinan di Kepulauan Nias," PhD Thesis, Institut Teknologi Sepuluh Nopember, 2015.
[23] A. Noviani, K. Fithriasari, and Irhamah, "Small area estimation with hierarchical bayesian neural network approach for case dropout children in poverty in East Java Province," PhD Thesis, Institut Teknologi Sepuluh Nopember, 2016.
[24] R. N. Yani, F. Yanuar, and H. Yozza, "Inferensi bayesian untuk σ2 dari distribusi normal dengan berbagai distribusi prior," *J. Mat. UNAND*, vol. 7, pp. 132–139, 2018.

[25] Azizah, "Pemodelan klaim asuransi menggunakan pendekatan bayesian dan markov chain monte carlo," *J. Kaji. Mat. Dan Apl.*, vol. 2, pp. 7–13, July 2021.
[26] U. Destiarina, M. Hadijati, D. Komalasari, and N. Fitriyani, "Estimasi parameter distribusi mixture eksponensial dan weibull dengan metode bayesian markov chain monte carlo," *Eig. Math. J.*, vol. 2, pp. 28–38, June 2019, doi: 10.29303/emj.v1i1.30.
[27] J. Luo, L. D. Carolis, B. Zeng, and M. Jeon, "Bayesian estimation of latent space item response models with jags, stan, and nimble in R," *Psych*, vol. 5, no. 2, pp. 396–415, May 2023, doi: 10.3390/psych5020027.
[28] M. Hecht, S. Weirich, and S. Zitzmann, "Comparing the mcmc efficiency of jags and stan for the multi-level intercept-only model in the covariance- and mean-based and classic parametrization," *Psych*, vol. 3, no. 4, pp. 751–779, Nov. 2021, doi: 10.3390/psych3040048.
[29] C. C. Monnahan, J. T. Thorson, and T. A. Branch, "Faster estimation of bayesian models in ecology using hamiltonian monte carlo," *Methods Ecol. Evol.*, vol. 8, no. 3, pp. 339–348, Mar. 2017, doi: 10.1111/2041-210X.12681.
[30] B. Liu, "Hierarchical bayes estimation and empirical best prediction of small-area proportions," PhD Thesis, College Park, University of Maryland, 2009.
[31] I. Ayuningtyas, "Small Area Estimation Pada Kasus Respon Multinomial dengan Pendekatan Hierarchical Bayes (Aplikasi pada Proporsi Pengangguran menurut Kategori Pengangguran di Pulau Kalimantan, 2015)," PhD Thesis, Institusi Teknologi Sepuluh Nopember, 2017.