



# Small Area Estimation of Multidimensional Poverty in East Java Province Using Satellite Imagery

H C L A Ridho<sup>1,\*</sup>, R B Prasetyo<sup>1</sup>

<sup>1</sup> Politeknik of Statistik STIS, Jl. Otto Iskandardinata No.64C, Jakarta, Indonesia

\* Corresponding author's e-mail: [helencantikalauraar@gmail.com](mailto:helencantikalauraar@gmail.com)

**Abstract.** The government has so far focused on a monetary approach to overcoming poverty, while poverty is multidimensional. Holistic and accurate poverty indicators are needed as material for policy formulation, such as the Multidimensional Poverty Index (IKM), which is calculated from raw data from the National Socioeconomic Survey (SUSENAS). However, the direct estimation of the multidimensional poverty headcount (AKM) is only accurate at the provincial level, as seen from the relative standard error (RSE) of several districts and cities, which is still above 25 percent. Increasing the sample size requires time, effort, and cost, so the Small Area Estimation (SAE) method can be an alternative. Apart from using official statistics for accompanying variables, satellite imagery has the advantage of being up-to-date and available up to a granular level. This study aims to estimate the AKM at the district/city level in East Java Province by utilizing satellite imagery and official statistics in SAE. The results showed that SAE HB Beta-logistics, with the accompanying variables combined with satellite imagery and official statistics, has a higher accuracy than direct estimation.

## 1. Introduction

In tackling poverty at a macro level, the poverty calculation commonly carried out by the Central Bureau of Statistics (BPS) uses an economic approach known as monetary poverty. Through this approach, poverty is seen as an economic inability to meet basic food and non-food needs based on expenditure. Based on the March National Social Economic Survey (SUSENAS), the level of monetary poverty in Indonesia has increased over the last three years due to the COVID-19 pandemic, where the highest was in 2021 at 10.14 percent.

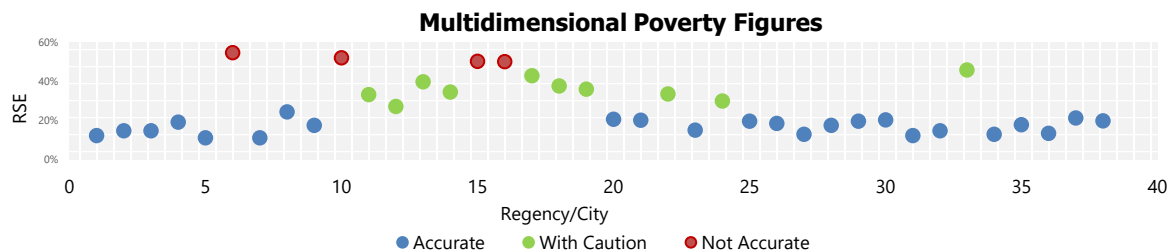
As the socio-economic transformation of society develops, perspectives regarding poverty are increasingly changing over time. The issue of poverty, if viewed from a monetary perspective alone, is not enough to describe the very complex problems in people's lives. Poverty needs to be measured by considering various social aspects in greater depth. Reviewing poverty from various perspectives (multidimensional) can provide information regarding deprivation or deficiencies that occur in one dimension. According to [1], if policy only focused on eliminating monetary poverty, the government would lose attention to basic social problems that are not directly related to income.

Since 2010, the United Nations Development Program (UNDP) and the Oxford Poverty and Human Development Initiative (OPHI) have agreed on a new approach to measuring poverty, namely the Multidimensional Poverty Index (MPI) [2]. According to [3], MPI does not try to eliminate monetary poverty but rather provides a broader and more measurable view of reducing all aspects of poverty. MPI consists of the dimensions of health, education, and living standards. Apart from that, this index has



been used by many countries to measure poverty levels in a more holistic way. Some indicators that are not available can be modified to suit the circumstances occurring in each country. In Indonesia, MPI is better known as the Multidimensional Poverty Index (IKM). There are two components that make up IKM, namely the multidimensional poverty headcount (AKM) and multidimensional poverty intensity, where IKM is the product of these two components. According to [4], AKM is the percentage of multidimensional poor people or people who are deprived in the dimensions of health, education, and living standards in the total population.

In realizing the achievements of the Sustainable Development Goals (SDGs) in the first goal, eliminating poverty in any form and anywhere, accurate and specific poverty data is needed down to the smallest level as a reference for policymakers. AKM's direct estimates using SUSENAS data for the March period are only accurate up to the provincial level. According to BPS (2019), estimation results are said to be accurate when they have a Relative Standard Error (RSE) value below 25 percent, can be used with caution when the RSE value is between 25 and 50 percent, and are inaccurate or unreliable when the RSE is above 50 percent. The RSE for direct estimates of AKM and population deprivation at the provincial level in East Java Province has shown a value of 3.55 percent, which is below 25 percent, and it can be said that the estimate is accurate. However, the RSE produced at the district or city level exceeds 25 percent and even 50 percent in some districts and cities.



**Figure 1.** RSE direct estimate of AKM at district/city level

In order to improve the precision of these estimates, it is necessary to add additional samples to districts or cities that have an RSE of more than 25 percent. However, this requires money, time, and energy. Indirect estimation methods, such as Small Area Estimation (SAE), can be a more efficient and accurate alternative to obtaining estimates for small areas by utilizing accompanying variables or additional information from neighboring sample units and regional characteristics that are known globally. The accompanying variables used in SAE are data obtained at the desired area level and must be related to the variables being estimated. Therefore, data whose availability meets the criteria is needed so that it can be relied on as an accompanying variable.

Many studies have used satellite imagery as a companion variable in SAE to estimate poverty levels. Research conducted by [6] estimating the average per capita expenditure at sub-district level in Yogyakarta Special Region Province using satellite imagery in the SAE EBLUP Fay-Herriot shows that the smallest MSE and RSE are produced in the model with the accompanying variables of a combination of satellite imagery and the Village Potential Statistics (PODES). Satellite image data used in this research includes build-up area (BUI), land surface temperature (LST), nighttime light intensity (NTL), and air pollution. Then [7] also conducted research related to sub-district level poverty estimates in West Java Province using SAE EBLUP Fay-Herriot with the use of satellite imagery and accompanying variables such as NTL, LST, CO emissions, NDVI, and NDBI. Based on the research results, it was found that the satellite imagery accompanying variables produced the best estimates compared to other accompanying variable models.

However, research that uses satellite imagery to estimate poverty mostly uses SAE with the EBLUP approach, while SAE EBLUP assumes a linear mixed model that is suitable for continuous and normally distributed variables. In this study, the distribution of direct estimates of AKM and the deprived population tended to be skewed to the right, so there was an indication of a violation of the normality assumption. The hierarchical Bayes (HB) approach can be a solution because it can be applied to various



data distribution patterns, both in continuous variables as well as binary and count variables. The advantage of the HB method is flexibility in assuming a normal distribution and determining the distribution adaptively based on existing survey data [8]. The use of the HB beta-logistic model is based on the beta distribution, which has a range of 0 to 1 and can be calculated as the probability of a proportion or percentage of a population. Based on the comparison results of the SAE EBLUP and HB Beta-logistic methods on proportion data that is not normally distributed [9], it shows that SAE HB Beta-logistic provides more accurate results compared to SAE EBLUP and direct estimation.

Based on the results of related research, it can be concluded that the use of satellite imagery as an accompanying variable can help Podes provide more accurate estimates compared to direct estimates with limited samples. Apart from that, there has been no research that utilizes satellite imagery as a supporting variable for Podes data in SAE HB Beta-logistics to estimate AKM and the deprived population. East Java Province is the province with the highest number of monetary poor people in the last seven years. In 2021, the percentage of the financially poor population will reach 11.4 percent, or the equivalent of 4.57 million people. So, the objectives of this research are (1) to provide an overview of multidimensional poverty in East Java Province in 2021; (2) to estimate AKM and the percentage of deprived population at district or city level in East Java Province in 2021 using SAE HB Beta-logistic with accompanying variables a combination of satellite and Podes imagery; and (3) to visualize the estimated results of AKM and the deprived population at district or city level in East Java Province in 2021.

## 4. Methodology

### 2.1. Theoretical basis

**2.1.1. Multidimensional Poverty.** Sen (1999) defines multidimensional poverty as a measurement of poverty based on a lack of access to various opportunities and freedoms, such as access to health and education services, access to employment and capital, as well as access to fair decision-making and social participation. A person is said to be multidimensionally poor if he experiences deprivation (lack of poverty) indicators that are actually experienced by him. Because he is affected by deficiencies in various things, he is unable to achieve the things he can achieve because of these limitations [10].

The following table presents the dimensions and indicators of MPI used in the research. In this research, the selection of indicators is based on the availability of data, and the weights used have the same value for each indicator.

**Table 1.** Dimensions and indicators used in MPI

Dimensions	Indicator	Thresholds	Weigher
Health	Sanitation	Households that do not have public, shared, or private defecation facilities and the type of toilet is not a gooseneck	1/9
	Clean water	Households that consume clean water that does not come from protected wells, metered taps, retail taps, or protected springs and whose water source is < 10 meters away from the septic tank	1/9
	Calorie & Protein Consumption	Daily household calorie consumption is less than 70% of the RDA, and daily household protein consumption is less than 80% of the RDA.	1/9
Education	Preschool Education	Households that have preschool-age children (3-6 years) who do not have access to preschool services such as play groups, early childhood education (PAUD), kindergarten (TK), and other types of preschool education	1/9



Dimensions	Indicator	Thresholds	Weigher
Education	School Sustainability	Households that have elementary and middle school-age children who are unable to complete their education up to high school or equivalent	1/9
	Literacy	Households that have family members of productive age (15–64 years) unable to read and write	1/9
Standard of Living	Cooking Fuel	Households with types of fuel or energy for cooking consisting of kerosene, charcoal, briquettes, and firewood	1/9
	House Condition	Households that have at least inadequate conditions for two of the following three sub-indicators: 1. Roof: made other than concrete, tile, zinc and asbestos, such as bamboo, wood/shingle, straw/fiber/leaves/thatch and others 2. Walls: made from materials other than walls and wood, such as woven bamboo, wooden sticks, bamboo and other materials 3. Floors: made from materials other than marble, ceramic, granite, tiles, terrazzo, cement and wood, such as bamboo, low quality wood/boards, soil and other materials	1/9
	Home ownership	Households that do not have housing assets: contracts, rentals, rent-free belonging to other people, rent-free belonging to parents/relatives, and others	1/9

A person is identified as multidimensionally poor if his/her deprivation score ( $c_i$ ) is less than the poverty cutoff of 0.333.

$$c_i = w_1I_1 + w_2I_2 + \dots + w_dI_d \tag{1}$$

Where  $I_i = 1$  (if someone is deprived in indicator  $i$ ) and  $I_i = 0$  (if not deprived)  $w_i$  is the weight of indicator  $i$  with  $\sum_{i=1}^d w_i = 1$

$$H = \frac{q}{n} \tag{2}$$

$$A = \frac{\sum_{i=1}^n c_i(l)}{q} \tag{3}$$

$$IKM = H \times A \tag{4}$$

where  $H$  (multidimensional poverty headcount (AKM)) = proportion of the number of multidimensional poor people to the total population;  $A$  (multidimensional poverty intensity) = the weighted average proportion of indicators in which poor people are deprived;  $q =$  the number of individuals categorized as poor multidimensionally,  $n$  is the total population,  $l$  is the amount of deprivation that a person must experience to be categorized as poor, and  $c_i(l)$  is the deprivation sensor score.

**2.1.2. Small Area Estimation. Small Area Estimation (SAE)** is a statistical technique used to estimate population characteristics or parameters at the level of small areas with a relatively small sample size. This technique utilizes additional information from higher regional levels, such as broader survey results or census data, to increase the precision of estimates at lower regional levels [11]. SAE is an indirect estimation method that borrows strength from observations of adjacent area samples by utilizing additional information, in this case accompanying variables.

**2.1.3. Hierarchical Bayes (HB).** In the HB method, the prior distribution ( $f(\lambda)$ ) is determined by the researcher subjectively to obtain the posterior distribution ( $f(\theta | y)$ ) of the small area parameters that want to be estimated ( $\theta$ ) with the available data ( $y$ ). Inference in the HB method is based on the posterior distribution so that the parameter ( $\phi = h(\theta)$ ) is estimated by its posterior mean ( $\phi^{HB} = E[h(\theta)|y]$ ) and the precision measurement of the estimator is based on its posterior variance ( $V[h(\theta)|y]$ ). According



Rao & Molina (2015), to applying the Bayesian theorem, the posterior distribution is obtained through the integration process as follows:

$$f(\boldsymbol{\theta}|\mathbf{y}) = \int f(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{y})d\boldsymbol{\lambda} \quad (5)$$

$$f(\boldsymbol{\theta}|\mathbf{y}) = \int f(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\lambda})f(\boldsymbol{\lambda}|\mathbf{y}) d\boldsymbol{\lambda} \quad (6)$$

Where  $f(\boldsymbol{\theta} | \mathbf{y}, \boldsymbol{\lambda})$  is the distribution of  $\boldsymbol{\theta}$  considering  $\mathbf{y}$  and  $\boldsymbol{\lambda}$  and  $f(\boldsymbol{\lambda} | \mathbf{y})$  is the distribution of  $\boldsymbol{\lambda}$  considering  $\mathbf{y}$ . Posterior distribution and obtain its magnitude as the posterior average using the Markov Chain Monte Carlo (MCMC) integration method.

#### 2.1.4. SAE HB Beta-logistics

The use of the beta distribution was chosen to model proportion data because this distribution is suitable for data with a range of (0,1) with an asymmetric distribution [8]. If written in full, the beta - logistic model with unknown sampling variance can be written as follows:

$$1. \text{ sampling model} \\ \hat{\theta}_j | \theta_j, \boldsymbol{\beta}, \sigma_v^2 \sim \text{ind Beta}(a_j, b_j), j = 1, \dots, m \quad (7)$$

With  $a_j = \theta_{jk}$  and  $b_j = (1 - \theta_j)_k$  so  $E(\hat{\theta}_j) = \theta_j = \frac{a_j}{a_j + b_j}$ . Where  $\theta_j$  is the parameter to be estimated or the average proportion in the small area  $j$ ,  $a_j$  and  $b_j$  is a parameter of the beta distribution, and  $k$  is a constant which is assumed to be in the *gamma distribution* with the parameter  $g_1$  and  $g_2$ , i.e  $k \sim \text{Gamma}(g_1, g_2)$

#### 2. Linking models

$$\text{logit} \left( \frac{a_j}{a_j + b_j} \right) | \boldsymbol{\beta}, \sigma_v^2 \sim \text{ind } N(x_j^T \boldsymbol{\beta}, \sigma_v^2), j = 1, \dots, m \quad (8)$$

a tau ( $\tau$ ) can be written as:

$$\text{logit}(\theta_j) | \boldsymbol{\beta}, \sigma_v^2 \sim \text{ind } N(x_j^T \boldsymbol{\beta}, \sigma_v^2), j = 1, \dots, m \quad (9)$$

HB inference on parameters  $\theta_j$  assumes a flat prior for  $\boldsymbol{\beta}$  and  $\sigma_v^2$  which, if written in notational form, is  $\beta_j \sim N(\mu_{\beta_j}, \sigma_{\beta_j}^2)$  and  $\sigma_v^2 \sim IG(c_1, c_2)$ . *IG* is an abbreviation for Inverted Gamma. With  $\mu_{\beta_j}$  and  $\sigma_{\beta_j}^2$  are the prior mean and variance respectively  $\beta_j$ . Then  $c_1$  and  $c_2$  are the prior parameters of  $\sigma_v^2$ . In this case,  $g_1$ ,  $g_2$ ,  $\mu_{\beta_j}$ ,  $\sigma_{\beta_j}^2$ ,  $c_1$ , and  $c_2$  are made fixed or determined before the iteration process takes place (initial value). The beta-logistic model with unknown sampling variance can also be called SAE HB Beta-logistic.

#### 2.2. Data and Data Sources

In this research, the data used comes from three sources, namely Susenas Kor and Consumption & Expenditure Module for the period March 2021, Podes 2021 data, as well as data from satellite imagery for the observation period January 1 2021 to December 31 2021. The AKM and deprived population calculations use Susenas data, while accompanying variables in SAE modeling use data from processed satellite images and Podes. Satellite image data collection in this research uses the Python programming language with the Google Earth Engine Python API on Google Colab.

#### 2.3. Analysis Method

This research uses descriptive analysis and inferential analysis using software such as SPSS, R Studio, and QGIS. Descriptive analysis is presented through tables and graphs. Meanwhile, inferential analysis





was carried out to calculate direct estimates of AKM and the deprived population, as well as indirect estimates using the SAE HB beta-logistic method. Several stages of inferential analysis in this research are as follows:

### 2.3.1. Data Preparation Stage

1. The multidimensional poverty data preparation stage includes modifying indicators in the IKM that refer to the Alkire-Foster method, preparing IKM data from raw Susenas data for the March 2021 period, filtering the information needed for direct estimation, and forming new variables needed to calculate AKM and the deprived population.
2. The collection and preprocessing of remote sensing satellite image data is carried out via GEE. The preprocessing carried out includes cloud selection and cloud masking to get a collection of cloud-free images and median reducing to get the median value of the collected images, which represents satellite images from one year.
3. In Podes data, the data preparation stage is carried out by filtering the accompanying variables used in indirect estimation and then aggregating them according to district or city level.

2.3.2. *Direct Estimation Stage.* In this research, the Susenas March 2021 sampling design was used. Then exploration and normality testing were carried out on each variable of interest.

2.3.3. *Indirect Estimation Stage.* In this research, the 'saeHB' package is used, the indirect estimation stages can be described as follows.

1. Test the non-multicollinearity assumption on accompanying variables in the model by removing variables that have a correlation of more than 0.8. This aims to ensure that there are no accompanying variables that correlate with each other.
2. Forming the SAE HB Beta-logistic model by  $\hat{\theta}_i$  being a direct estimator of AKM and the percentage of the deprived population in the form of proportions. The number of updates, iterations, thins, and burn ins is chosen by trial and error until algorithm convergence is achieved. Updates are carried out to update prior information from the initial value based on the likelihood function or existing data. Thin is used to reduce autocorrelation between generated samples. Finally, iterate and burn in to remove the influence of the initial value.

2.3.4. *Model Evaluation Stage.* Evaluation is carried out by comparing the Relative Standard Error (RSE) of direct estimators and indirect estimators. RSE is a measure of the precision of an estimate relative to the estimate. The RSE value can be calculated using the following formula:

$$RSE(\hat{\theta}_i) = \frac{SE(\hat{\theta}_i)}{\hat{\theta}_i} \times 100\% \quad (10)$$

where  $SE(\hat{\theta}_i)$  = standard error of the response variable

2.3.5. *Visualization Stage.* The estimation results of all variables of interest with the best model are then visualized to obtain a multidimensional poverty map.

## 5. Result and Discussion

### 3.1. Multidimensional Poverty Profile of East Java Province

3.1.1. *Multidimensional and Monetary Poverty by Region.* According to regional classification, the number and percentage of poor people, both monetary and multidimensional, in rural areas is higher than in urban areas. According to World Bank (2021), rural areas often have limitations in terms of access to quality human resources, adequate infrastructure, and a diversified labor market. This results



in rural residents often relying on the agricultural sector which is vulnerable to climate change and market price fluctuations, making them more vulnerable to poverty.

**Table 2.** Comparison of poor and monetary population by region in East Java

Regional Classification	Number of Poor People (thousand souls)		Percentage of Poor Population (%)		Multidimensional Poverty Intensity	Multidimensional Poverty Index
	Monetary	Multidimensional	Monetary	Multidimensional		
Urban	1,840	920	8.38%	4.19%	36.09%	1.51%
Rural	2,733	2,107	15.05%	11.61%	36.94%	4.29%
Total	4,573	3,027	11.40%	7.55%	36.68%	2.77%

The multidimensional poor population is measured by considering various aspects, while the monetary poor population is measured only based on expenditure. In line with the findings [4], in this research, the number of multidimensional poor people in East Java in 2021 is lower than the number of monetary poor people, this may indicate that multidimensional poor people may have economic limitations, they can still survive and meet basic needs. by utilizing available resources.

*3.1.2. Deprivation by Dimensions and Indicators by Region.* A comparison of the percentage of deprived population in each dimension between urban and rural areas in East Java Province in 2021 can be seen in the following table.

**Table 3.** Deprivation in each dimension of MPI by region

Region	Health Dimensions	Dimensions of Education	Dimensions in Living Standards
Urban	1.80%	0.66%	0.57%
Rural	5.30%	2.06%	1.36%
Total	5.31%	2.05%	1.13%

There are similar types of contributors to multidimensional poverty between urban and rural areas, but with different compositions. Both in urban and rural areas, residents are deprived in the health dimension. This finding is in line with research [4], which explains that the health dimension is a major contributor to multidimensional poverty. Next, followed by the education dimension and the quality of life standard dimension.

**Table 4.** Deprivation in each MPI indicator by region

Region	Health Dimensions			Educational Dimensions			Living Standard Dimensions		
	Proper Sanitation	Drinking water	Calorie & Protein Consumption	Preschool Education	School Sustainability	Literacy	Cooking Fuel	House Condition	Home Ownership Status
Urban	6.87%	15.56%	8.93%	9.48%	2.29%	9.31%	5.96%	0.67%	14.63%
Rural	18.52%	22.46%	6.96%	8.72%	4.37%	23.28%	21.42%	2.07%	4.52%
Total	12.15%	18.68%	8.04%	9.14%	3.23%	15.63%	12.96%	1.30%	10.06%

The highest contributor to multidimensional poverty is based on indicators in each dimension, namely drinking water, literacy and cooking fuel. In the health dimension, residents in rural areas tend to be more deprived in the indicators of drinking water and adequate sanitation, while in urban areas they are more deprived in the indicators of calorie and protein consumption. Then in the education dimension, residents in rural areas tend to be more deprived in terms of literacy and school continuity indicators, while in urban areas they are more deprived in terms of preschool education indicators. In the standard of living dimension, residents in rural areas tend to be more deprived on indicators of cooking fuel and housing conditions, while those in urban areas are more deprived on indicators of home ownership status.



3.2. Direct Estimate

Data distribution can describe the distribution of data over a certain range of values and provide information about *central tendency*. Below is presented the distribution of data for each variable of interest.

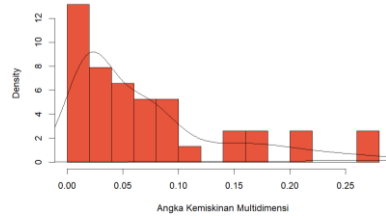


Figure 1. Distribution of direct estimator data

In Figure 1 it can be seen that the data distribution for all interest variables has a tendency to skew to the right. This shows that most of the direct estimator values are below the average value of the data. In addition, checking the normality assumption is necessary to decide on the use of the SAE HB method. Below are the results of testing the suitability of the distribution of direct estimators using the Shapiro Wilk test.

Table 5. Normality Test Results (*Shapiro Wilk*)

Interest Variable	Test Statistics (W)	P-value	Information
Multidimensional Poverty Figures	0.8365	0.0000	Not normally distributed

Based on the results of the normality test, it can be concluded that all interest variables are not normally distributed. Using HB in SAE can be more flexible to adjust data distribution, choosing a beta distribution is suitable to accommodate differences. Therefore, this study uses SAE HB Beta-logistic to estimate the four variables of interest.

3.3. Indirect Estimate (*SAE HB Beta-logistic*)

3.3.1. Selection of Concomitant Variables. Below are presented the results of the Pearson correlation test on several candidate accompanying variables.

Table 6. Pearson Correlation Test Results

Variable	Correlation	P-value	Information
Hospital Ratio ( $X_1$ )	-0.3628	0.0252	Significant
Maternity Hospital Ratio ( $X_2$ )	-0.0893	0.5940	Not significant
Inpatient Health Center Ratio ( $X_3$ )	0.1564	0.3483	Not significant
Polyclinic Ratio ( $X_4$ )	-0.4547	0.0041	Significant
Polides Ratio a ( $X_5$ )	-0.4049	0.0117	Significant
Pharmacy Ratio ( $X_6$ )	-0.3925	0.0148	Significant
Hospital to Polyclinic Ratio ( $X_7$ )	-0.4480	0.0048	Significant
Kindergarten Ratio ( $X_8$ )	-0.1475	0.3769	Not significant
High School Ratio ( $X_9$ )	0.1627	0.3292	Not significant
Junior High School Ratio ( $X_{10}$ )	0.2271	0.1703	Not significant
Senior High School ratio ( $X_{11}$ )	0.0976	0.5599	Not significant
College Ratio ( $X_{12}$ )	-0.3158	0.0535	Significant
Education Facilities Ratio ( $X_{13}$ )	0.1639	0.3256	Not significant
Average Years of Schooling ( $X_{14}$ )	-0.6821	0.0000	Significant
Bank Ratio ( $X_{15}$ )	-0.3346	0.0401	Significant
Minimarket Ratio ( $X_{16}$ )	-0.3704	0.0221	Significant
Restaurant Ratio ( $X_{17}$ )	-0.3728	0.0212	Significant

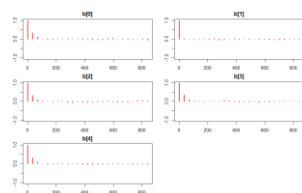




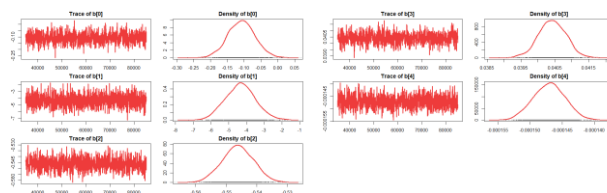
Variable	Correlation	P-value	Information
Bank and Restaurant Ratio ( $X_{18}$ )	-0.3766	0.0198	Significant
Ratio of Households Working in the Agricultural Sector ( $X_{19}$ )	0.3908	0.0153	Significant
Ratio of Household Electricity Users ( $X_{20}$ )	-0.0905	0.5889	Not significant
Population Density ( $X_{21}$ )	-0.4860	0.0020	Significant
Working Population Ratio ( $X_{22}$ )	0.4519	0.0044	Significant
PKH Beneficiary Families ( $X_{23}$ )	0.3179	0.0518	Significant
Regency/City Minimum Wage ( $X_{24}$ )	-0.4317	0.0068	Significant
NTL ( $X_{25}$ )	-0.4142	0.0097	Significant
Percentage of Green Open Space ( $X_{26}$ )	-0.3233	0.0477	Significant
LST ( $X_{27}$ )	-0.4943	0.0016	Significant
NDVI ( $X_{28}$ )	0.5539	0.0003	Significant
NDWI ( $X_{29}$ )	0.4236	0.0080	Significant
CO emissions ( $X_{30}$ )	0.0063	0.9700	Not significant
NO2 emissions ( $X_{31}$ )	-0.5066	0.0012	Significant
SO2 emissions ( $X_{32}$ )	-0.2775	0.0917	Significant

After selecting variables that have a significant correlation, a non-multicollinearity test is then carried out on each model by calculating the VIF (Variance Inflation Factor) value for each variable. A VIF value of more than 10 indicates a violation of the multicollinearity assumption, so this variable must be deleted. The selected accompanying variables and their VIF values are  $X_4(1.507)$ ,  $X_{14}(2.193)$ ,  $X_{22}(2.877)$ , and  $X_{27}(1.500)$ .

**3.3.2. Model Convergence.** After obtaining the selected variables in each model, the next step is to ensure that the model has converged or achieved stability in the calculations. In the model convergence process, iteration or repetition of calculations is carried out until the estimated values obtained do not change significantly. Bayesian inference that uses the MCMC technique to obtain the posterior distribution. In the AKM model estimation, the MCMC iterations used were 85,000, updated 25 times, thin 35, and burn in 35,000. The convergence of the MCMC algorithm can be detected through several diagnostic plots as follows:



**Figure 2.** Autocorrelation plot in the AKM model



**Figure 3.** Trace plot and density plot in the AKM model

Based on Figure 2, no periodic pattern was found in the trace plot and the density plot (which resembles a bell curve tends to be smooth for all parameters in the entire model). Then the autocorrelation plot also shows that all parameters in each are cut off after the first lag. This shows that there is no correlation between the parameter values generated and the convergence process is fast [9]. So, it can be concluded that the MCMC algorithm has converged.

**3.3.3. SAE HB Beta-Logistic Model Coefficient Estimation.** Below is a complete presentation of the estimated parameter coefficients for the SAE HB Beta-logistic model along with the 95% credible interval in the following table.

**Table 7.** Estimated SAE HB Beta-logistic model parameter coefficients along with 95% credible intervals

Parameter Coefficient	Mean	Standard Deviation	2.5%	97.5%
$\beta_0$	-0.1075780593	$3.998065 \times 10^{-2}$	-0.187574552	-0.1075141195
$\beta_1$	-4.3321808205	$8.458831 \times 10^{-1}$	-5.940757216	-4.3328212260
$\beta_2$	-0.5461346425	$5.071647 \times 10^{-3}$	-0.556023246	-0.5461929143
$\beta_3$	0.0404242854	$4.018561 \times 10^{-4}$	0.039626171	0.0404272623
$\beta_4$	-0.0001469871	$2.447031 \times 10^{-6}$	-0.000151727	-0.0001469514

A parameter is said to be significant in the model if its credible interval does not contain a zero value. It can be concluded that all accompanying variables significantly influence the logit of the direct estimates of the four variables of interest because the 95% credible interval does not exceed zero (see Table 7). In the AKM  $\beta_3$  estimation model, it is obtained that the estimated parameter coefficients m has a positive value, while the parameters  $\beta_1$ ,  $\beta_2$ , and  $\beta_4$  m has a negative value.

**3.3.4. Estimation results with SAE HB Beta-logistic.** In general, the following table presents a summary of the estimation results from SAE HB Beta-logistic and it is found that the distribution of AKM estimation results using SAE HB Beta-logistic follows the direct estimate results.

**Table 8.** Summary of SAE HB Beta-logistic estimation results

Model	Min	Q1	Median	Mean	Q3	Max
Direct Estimate	0.0053	0.0197	0.0558	0.0757	0.0931	0.2625
SAE HB Beta-logistics	0.0055	0.0200	0.0559	0.0757	0.0929	0.2621

**3.3.5. Model Evaluation.** After obtaining the estimation results, an evaluation process needs to be carried out to prove that the SAE model is able to produce estimators that are more accurate and precise than direct estimators. The evaluation measure is calculated using the RSE value of each estimation model.

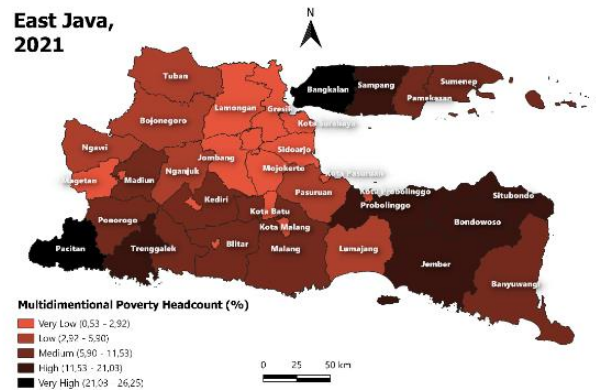
**Table 9.** Summary of RSE direct estimate and SAE HB Beta-logistic

Model	Min	Q1	Median	Mean	Q3	Max
Direct Estimate	10.80	14.92	19.71	25.52	34.62	54.35
SAE HB Beta-logistics	2.31	4.21	7.87	9.74	13.06	27.57

Based on Table 9, in general it can be seen that there has been a decrease in the RSE value estimated directly for SAE HB Beta-logistic. This shows that multidimensional poverty estimation using SAE HB Beta-logistic can improve the accuracy of direct estimates.

### 3.4. Multidimensional Poverty Visualization by Regency/City

Based on the AKM estimation results at the district/city level in East Java Province in 2021, visualization was then carried out using thematic maps to provide a spatial picture of multidimensional poverty. Poverty indicator data is presented in the form of a thematic map. It is hoped that the data and information presented in the form of thematic maps will make it easier for data users to make comparisons of poverty indicators between districts/cities in East Java Province. The thematic maps are all grouped into five categories using natural breaks.



**Figure 4.** SAE HB Beta-logistic estimation results by district/city

Figure 4 shows the regional distribution according to AKM based on five categories. There are two areas that fall into the very high multidimensional poverty category, namely Bangkalan Regency and Pacitan Regency. Then in the very low multidimensional poverty category, namely Magetan Regency, Blitar City, Kediri City, Blitar City, Malang City, Gresik Regency, Lamongan Regency, Sidoarjo Regency, Surabaya City and Pasuruan City.

## 6. Conclusion and Recommendation

Based on the results and discussion, there are several conclusions in this research.

1. In general, in East Java Province in 2021, most of the multidimensional poor population is in rural areas. According to the dimensions, the highest percentage of the deprived population is in the health dimension (5.31%) and the lowest is in the standard of living dimension (1.13%). According to the indicators, the indicators that have the largest percentage of deprivation include drinking water (18.68%), literacy (15.63%), and cooking fuel (12.96%), while the lowest is housing conditions (1, 30%), school sustainability (3.23%), calorie and protein consumption (8.04%).
2. The overall SAE HB Beta-logistic model with satellite and village imagery accompanying variables in estimating AKM shows a lower RSE value compared to the RSE of direct estimation. This proves that the SAE method can increase the precision of direct estimates and can estimate unsampled areas.

The recommendations that can be given to the East Java Provincial government and further research are as follows.

1. For the East Java Provincial government, apart from using monetary poverty calculations, it is necessary to consider a multidimensional approach where the estimation can use the indirect estimation method (SAE). Apart from that, Bangkalan Regency as the district with the highest multidimensional poverty rate can be a concern in policy making.
2. The multidimensional poverty estimates in this study are still at the district/city level, so it can be recommended in future research to estimate at a smaller level such as the sub-district level by exploring accompanying variables from other big data sources.

## References

- [1] A. Sen, *Development as freedom*. Oxford Paperbacks, 1999.
- [2] UNDP (United Nations Development Programme), "Human Development Report 2010," *UNDP (United Nations Development Programme)*, 2010.
- [3] S. Alkire and M. E. Santos, "A Multidimensional Approach: Poverty Measurement & Beyond," *Soc Indic Res*, vol. 112, no. 2, pp. 239–257, Jun. 2013, doi: 10.1007/s11205-013-0257-3.
- [4] C. Nurul *et al.*, *INDEKS KEMISKINAN MULTIDIMENSI INDONESIA 2015-2018*. 2020.



- [5] Badan Pusat Statistik, “Sampling Error Estimasi Hasil Survei Struktur Ongkos Usaha Tanaman Hortikultura 2018,” Nov. 2019. Accessed: Jun. 10, 2023. [Online]. Available: <https://www.bps.go.id/publication/2019/11/27/bf81027e6ab5b10666328e69/sampling-error-estimasi-hasil-survei-struktur-ongkos-usaha-tanaman-hortikultura-2018.html>
- [6] S. M. Hakim, “Penerapan Big Data pada Small Area Estimation Pengeluaran per Kapita (Studi Kasus : Provinsi Daerah Istimewa Yogyakarta),” Politeknik Statistika STIS, Jakarta, 2021.
- [7] S. D. Maretna, “Ekstraksi Citra Satelit Sebagai Variabel Penyerta SAE Tingkat Kemiskinan pada Level Kecamatan (Studi Kasus: Sembilan Kota di Provinsi Jawa Barat),” Politeknik Statistika STIS, Jakarta, 2022.
- [8] B. Liu, “Hierarchical Bayes Estimation and Empirical Best Prediction of Small-Area Proportions [Dissertation],” University of Maryland, College Park, 2009.
- [9] N. P. Istiqomah and I. Y. Wulansari, “Estimasi Angka Partisipasi Kasar Perguruan Tinggi Level Kabupaten/Kota di Pulau Kalimantan Tahun 2020 dengan Small Area Estimation Hierarchical Bayes Beta-Logistic,” *Seminar Nasional Official Statistics*, pp. 137–146, 2022.
- [10] BSD Bappeda DIY, “Laporan Akhir Analisis Kriteria dan Indikator Kemiskinan Multidimensi untuk Diagnostik Kemajuan Daerah di Daerah Istimewa Yogyakarta,” 2017.
- [11] J. N. K. Rao, *Small Area Estimation*, 1st ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003. doi: 10.1002/0471722189.
- [12] J. N. K. Rao and I. Molina, *Small Area Estimation*. Wiley, 2015. doi: 10.1002/9781118735855.
- [13] World Bank, “Poverty, median incomes, and inequality in 2021: a diverging recovery 1.” 2021. Accessed: May 14, 2023. [Online]. Available: <https://documents1.worldbank.org/curated/en/936001635880885713/pdf/Poverty-Median-Incomes-and-Inequality-in-2021-A-Diverging-Recovery.pdf>