DOI: https://doi.org/10.34123/jurnalasks.v16i2.799

POLITEKNIK STATISTIKA STIS

Jurnal Aplikasi Statistika & Komputasi Statistik
Vol. 16 No. 2, June 2024
e- ISSN: 2615-1367, p-ISSN:2086-4132

Available online at https://jurnal.stis.ac.id/

Small Area Estimation Approaches Using Satellite Imageries Auxiliary Data for Estimating Per Capita Expenditure in West Java, Indonesia

Muhamad Feriyanto[1], Arie Wahyu Wijayanto[2*], Ika Yuni Wulansari[3], Novia Budi Parwanto[4]

[1]BPS-Statistics Barru Regency, Indonesia, [2]Politeknik Statistika STIS, Jakarta, Indonesia, [3]University of Technology Sydney, Sydney, Australia, [4]IsDB Group Chief Economist, Islamic Development Bank, Jeddah, Saudi Arabia
**Corresponding Author: E-mail address:* ariewahyu@stis.ac.id

ARTICLE INFO

Article history:
Received 03 September, 2024
Revised 09 December, 2024
Accepted 09 December, 2024
Published 31 December, 2024

Keywords:
EBLUP; Expenditure per Capita; SAE; Remote Sensing; NTL

Abstract

Introduction/Main Objectives: The economy of a country can determine the welfare of its people. One of the economic indicators in Indonesia is per capita expenditure, which has the lowest estimation at the district level. **Background Problems:** Sub-district level estimates provide detailed information on inequality that cannot be explained at the district level. Unfortunately, sub-district level estimates of per capita expenditure in Indonesia have poor Relative Standard Error (RSE) values. **Research Method**: The Small Area Estimation (SAE) method can improve estimator accuracy on small samples by using auxiliary variable information. **Novelty:** The existence of big geospatial data such as remote sensing provides an advantage in the efficient use of auxiliary variables. **Finding Result**: The Empirical Best Linear Unbiased Prediction (EBLUP) model using Nighttime Light Intensity (NTL) as an auxiliary variable provides the best results of the five proposed models. Remote sensing data can potentially be used in SAE auxiliary variables. It provides opportunities for cheaper, faster, and more efficient data collection compared to conventional data.

1. Introduction

The economy of both developed and emerging nations is vitally important. Each individual has a significant part in the expansion and progress of a nation's economy. As social beings, individuals generally participate in collective transactions such as buying and selling. The economic volatility of a country is predominantly determined by its purchasing and selling operations. Undoubtedly, each household allocates a portion of its earnings towards meeting its daily necessities, including but not limited to education, healthcare, sustenance, taxes, and real estate. Consumers and producers are two economic categories that households might occupy. In addition to providing goods and services, households may also provide capital, land, and entrepreneurialism [1].

As per the classification provided by the BPS-Statistics Indonesia (Badan Pusat Statistik, BPS) [1], food expenditure and non-food expenditure. Food expenditure refers to the monthly amount that each household spends on food to meet its total consumption. Whereas, expenditure for non-food is the total consumption each household spends in a month for non-food, such as education, health, tax, etc. At now, the estimation of per capita expenditure in Indonesia is conducted at the district level, the lowest attainable level. Per capita gross domestic product in Indonesia amounted to IDR 56 million in 2018, which is equivalent to IDR 4.67 million per month. This economic value increased by 5.17 percent compared to the previous year [2]. Provinces in Java contributed the most to this economic value expansion, which also surpassed the target.

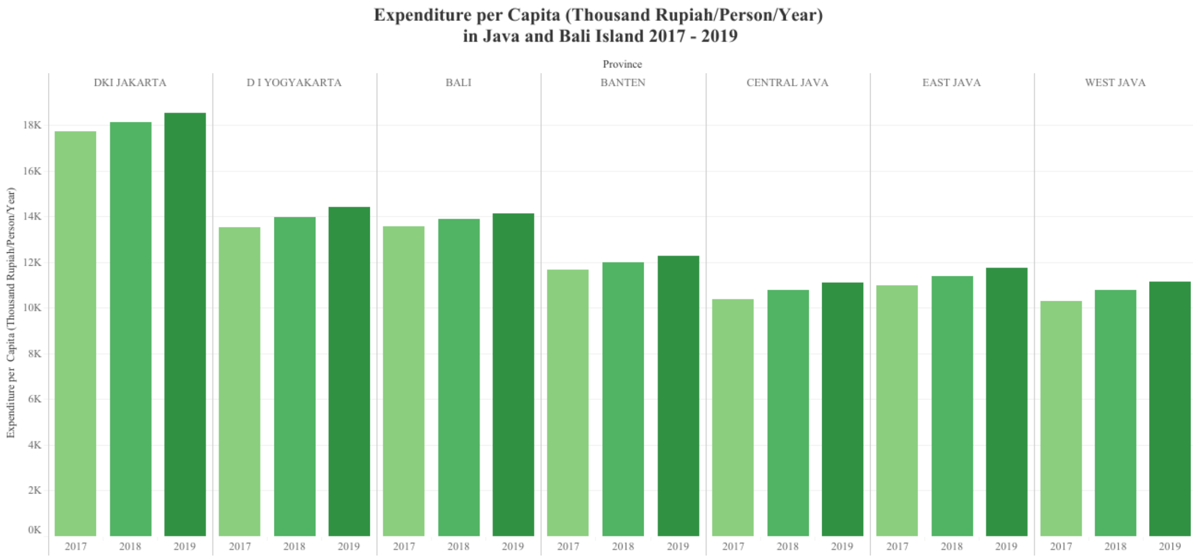


Figure 1. Expenditure per Capita in Java and Bali Island 2017-2019

West Java, situated on Java Island, is the second most populated province in Indonesia, trailing only Jakarta in terms of population [3]. Despite being one of the provinces on Java Island—the administrative heart of Indonesia—West Java ranked fifteenth in Indonesia in terms of per capita expenditure in 2018. Figure 1 shows that West Java ranks lowest out of all the provinces in Java and Bali Island from three years, 2017 until 2019. The province on Java Island should, in theory, be more populous and reflect higher levels of human well-being as depicted in Figure 1 on Expenditure per Capita in Java and Bali Island 2017-2019. As it ought to, the province on the island of Java is capable of accommodating a greater populace and enhancing human welfare.

Estimation of expenditure per capita at the district level only reflects the region as a whole, whereas, in reality, the situation in the district is not homogeneous. The estimated value of per capita expenditure at a lower level (such as the sub-district level) can be used by stakeholders as the foundation for good policy-making. The lower the level, the more detailed the information can be obtained. Unfortunately, per capita expenditure in Indonesia is only assessed at the district level since the lower the administrative level, the lower the estimated value, as demonstrated by the high Relative Standard Error (RSE) value. In Indonesia, a good estimated value is one with an RSE value of less than 25% [4]. Estimation of expenditure per capita at a lower level, which has a minimum sample, can be done using the Small Area Estimation (SAE) model.

Small Area Estimation (SAE) is an indirect estimation method that borrows the strength of auxiliary variables to gain the variation of some variables from the area between [5]. According to Rao and Molina (2015), an explicit model of Small Area Estimation is a model that considers the variability of the area between and combines the linear mixed model and generalized linear mixed model. The SAE model can be classified into two, unit level and area level. Area-level models are often used in various analyses because the ease of data in the level area can be obtained. The feasibility of Small Area Estimation model is determined from the goodness of auxiliary variables used.

By reflecting or sending waves to Earth, a technology known as remote sensing can be used to gather information about the planet [6]. Satellite imagery is the end product of remote sensing, and it contains information that can be extracted. Satellite imagery is the end product of remote sensing, and it contains information that can be extracted. For instance, The NOAA-VIIRS creates a product called Nighttime Light Intensity (NTL), which provides data on the index used to observe the level of electricity in the area [7]. Currently, the results of remote sensing have been used in various fields such as poverty [8], [9], [10], [11], [12], land cover [13], [14], estimation of electricity [7], [15], urban identification [16], and many more.

Nowadays, the application of the Small Area Estimation model is already in various fields, such as estimation in agriculture indicators [17], [18], [19], poverty estimation [20], [21], infant mortality rate (IMR) [22], food insecurity [23], [24], and expenditure per capita [25], [26]. Big data has also been widely employed in SAE modelling, in addition to data from surveys and censuses. One type of big data that is frequently utilized in Small Area Estimation modelling is remote sensing. In order to represent the whole region down to the smallest grid, auxiliary variables must have error-free coverage of the entire region. Similarly, satellite imagery findings are thought to cover the region extensively. For instance, Singh et al. (2002) modelled crop yield estimation using the Normalized Different Vegetation Index (NDVI) and the Ratio Vegetation Index (RVI), while Kaban et al. (2022) estimated spending per capita using Nighttime Light Intensity (NTL).

This study follows on from the prior research [25] on estimation per capita using Nighttime Light Intensity. The purpose of this study's renewal is to broaden the use of remote sensing for auxiliary variables in SAE. Furthermore, the data from satellite imagery, Nighttime Light Intensity (NTL) and Land Surface Temperature (LST) for auxiliary variables are combined in this study. The use of big data, particularly remote sensing in small area models, allows the government to reduce the Relative Standard Error (RSE), allowing it to make public policy in sub-districts or other level areas with a small sample size. Furthermore, the abundance and low cost of remote sensing data, as well as the comprehensive coverage of remote sensing data, means that estimates can be made more efficiently, both in terms of cost and time. Hence, a combination of remote sensing and Small Area Estimation provides the estimation expenditure per capita in granular or minimum sample areas, such as sub-districts more effectively and efficiently.

2. Material and Methods

2.1. Type of Research

This study assesses per capita expenditure at the sub-district area using quantitative approach. The method Empirical Best Linear Unbiased Predictor (EBLUP) is used to compare the proposed model with several possible auxiliary variables datasets.

2.2. Location and Time Research

Located in the west of Java Island near Banten and Central Java Provinces, West Java Province had 27 districts and 626 sub-districts in 2018. This is a cross-sectional study that makes use of secondary data from Statistics Indonesia (Badan Pusat Statistik, BPS) for conventional data and remote sensing data as the application of the usage of big data. The Indonesian National Socio-Economic Survey (SUSENAS) in March 2018 and Village Potential (PODES) 2018 data were used in this study. SUSENAS is a biannual survey conducted by BPS every March and September. Whereas, PODES is a census conducted by BPS three times every ten years, exactly one year after the main census. Data remote sensing was collected in the Google Earth Engine (GEE) catalogue which provides many sources of remote sensing data from many types of satellites.

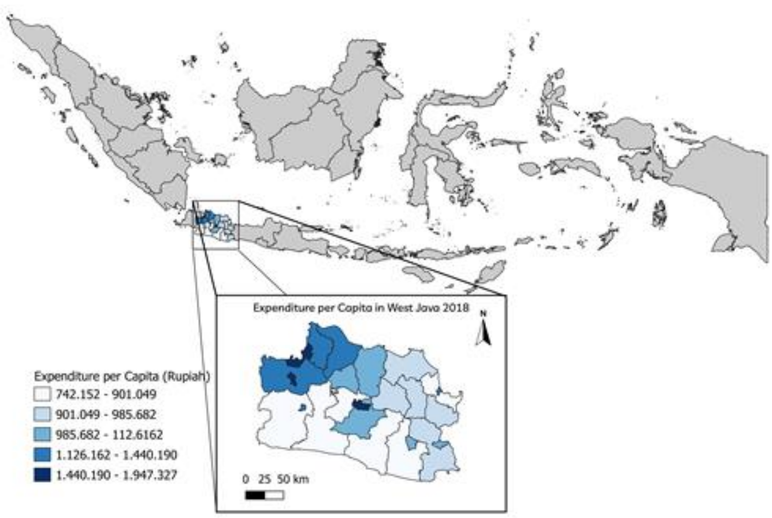


Figure 2. West Java (Area of Research Interest)

2.3. Research Variables

Expenditure per capita was the main estimator collected in SUSENAS in March 2018. Per capita expenditure in SUSENAS is calculated by adding the sum of food and non-food expenditures in a month from the consumption and expenditure questionnaire. The variables obtained by PODES, which are detailed in Table 1, are included as auxiliary variables in the Small Area Estimation model to model per capita expenditure. All variables in Table 1 are aggregated to the sub-district level, beginning with raw data collected at the village and household levels, to facilitate estimations at the sub-district level.

Table 1. Conventional Data

Symbol	Variable	Description	Source
Y	Expenditure per Capita	Expenditure per capita in a month in Indonesian Rupiah (IDR)	SUSENAS 2018
$X1$	University	Total of University	PODES 2018
$X2$	Islamic Boarding School	Total of Islamic Boarding School	PODES 2018
$X3$	Mechanical Training	Total of Mechanical Training Place	PODES 2018
$X4$	Language Training	Total of Language Training Place	PODES 2018
$X5$	Medical Facility	Total of Medical Facility	PODES 2018
$X6$	Mini Market	Total of Mini Market	PODES 2018

Source: Statistics Indonesia (Badan Pusat Statistik, BPS)

The application of remote sensing in this study uses two variables from remote sensing, Nighttime Light Intensity (NTL) and Land Surface Temperature (LST). The Asian Development Bank (2016), ADB used NTL data as an approach to socioeconomics indicators in several developing countries. The NTL data is strongly positively correlated with average household expenditure and GDP growth in developing countries [28]. The occurrence of economic growth can be followed by the population growth, which indicated by increasing luminosity of nighttime light intensity. The increase in population in urban areas has a tendency to increase land surface temperature as one of the effects of urban heat islands [29].

These two kind of variables are used individually and combined to get the best estimate. From Table 2, the NTL data in 2018 were obtained from the Suomi-NPP Satellite with Visible Infrared Imaging Radiometer Suite (VIIRS) instrument which has 750 meters of spatial resolution. Whereas LST data in 2018 were obtained from Terra Satellite with a Moderate Resolution Imaging Spectroradiometer (MODIS) instrument which has 1,000 meters of spatial resolution.

Table 2. Remote Sensing Data

Variable	Satellite	Resolution		Band used
		Spatial	Temporal	
Nighttime Light Intensity (NTL)	Suomi-NPP	750 m	16 days	avg_rad
Land Surface Temperature (LST)	Terra	1,000 m	16 days	LST_Day_1

This research uses remote sensing, a big data application, in conjunction with the statistical method. According to Figure 5, the analysis starts by obtaining raw data SUSENAS and PODES 2018 from BPS. Afterwards, the BPS-specified weighted value for design sampling was used for the calculation of subdistrict-level expenditure per capita. The formula used to calculate the expenditure per capita using weights is as follows:

$$\bar{y}_i = \frac{\sum_{i=1}^{n} w_{ij} y_{ij}}{\sum_{i=1}^{n} w_{ij}} \tag{1}$$

where $\bar{y}_i$ is average expenditure per capita in sub-district i^{th}, w_{ij} is a weighted factor in j^{th} household in i^{th} sub-district, and y_{ij} is expenditure per capita a month in j^{th} household in i^{th} sub-district. This estimation is known as direct estimation from subdistrict level expenditure per capita.

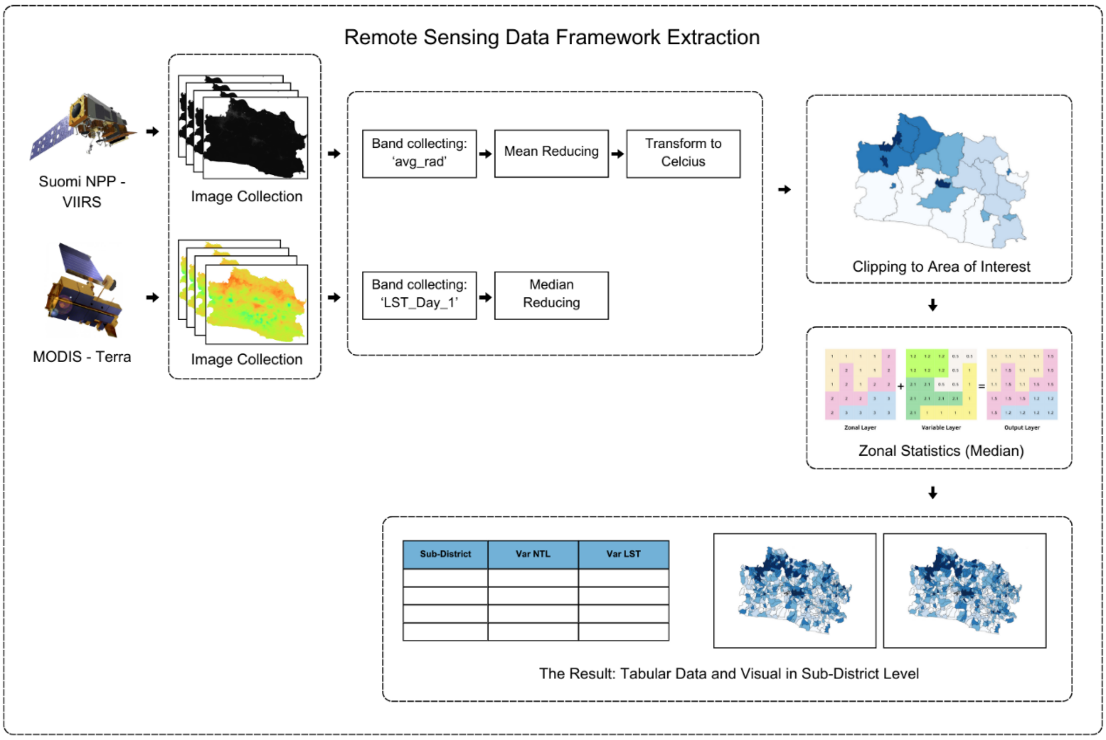


Figure 3. Remote Sensing Data Framework Extraction

All datasets are aggregated into sub-district levels. Data from SUSENAS is aggregated using formula (1), which uses unique weights for every household. PODES data obtained at the village level is aggregated by adding up all the total sub-district variables. Whereas, the remote sensing data was obtained from the image collection throughout 2018 from its satellite and its band. The image collection containing the band used in this study is reduced by mean or median reducer as shown in Figure 3. The result is single layer from NTL or LST in 2018 followed by clipping to area of interest (West Java) for zonal statistics. The zonal statistics concept is to take average or median of all pixel value of the variables within the administrative area (sub-district). Figure 4 shows that all of the variable values in the same zonal statistics administration will be averaged. Hence, the output layer represents the variable value (NTL or LST) in each sub-district.

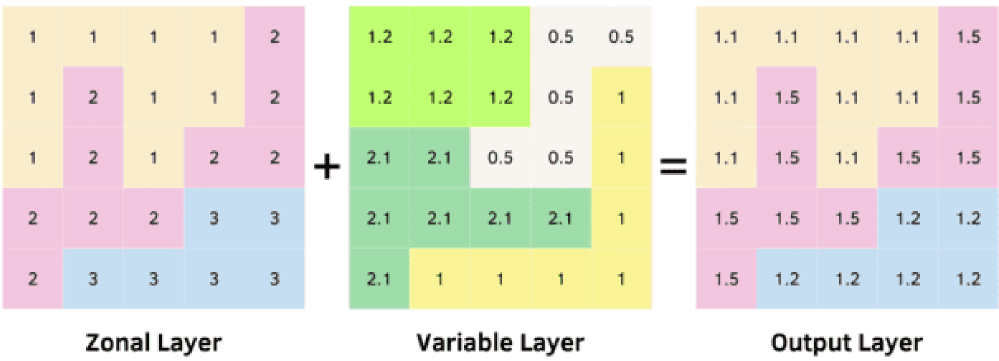


Figure 4. Zonal Statistics Illustration

Small Area Estimation (SAE) is one of the techniques of indirect estimation that borrows strength from auxiliary variables to estimate an estimator that has an inadequate sample [5]. SAE is typically divided into two distinct levels of estimation: area level and unit level. Due to the simplicity of data acquisition, area level is frequently used. Model Empirical Best Linear Unbiased Prediction (EBLUP) is used to estimate expenditure per capita in West Java. This model is an enhanced version of the preceding model BLUP; it estimates σ_v^2 using $\hat{\sigma}_v^2$ since that value σ_v^2 is unknown in reality.

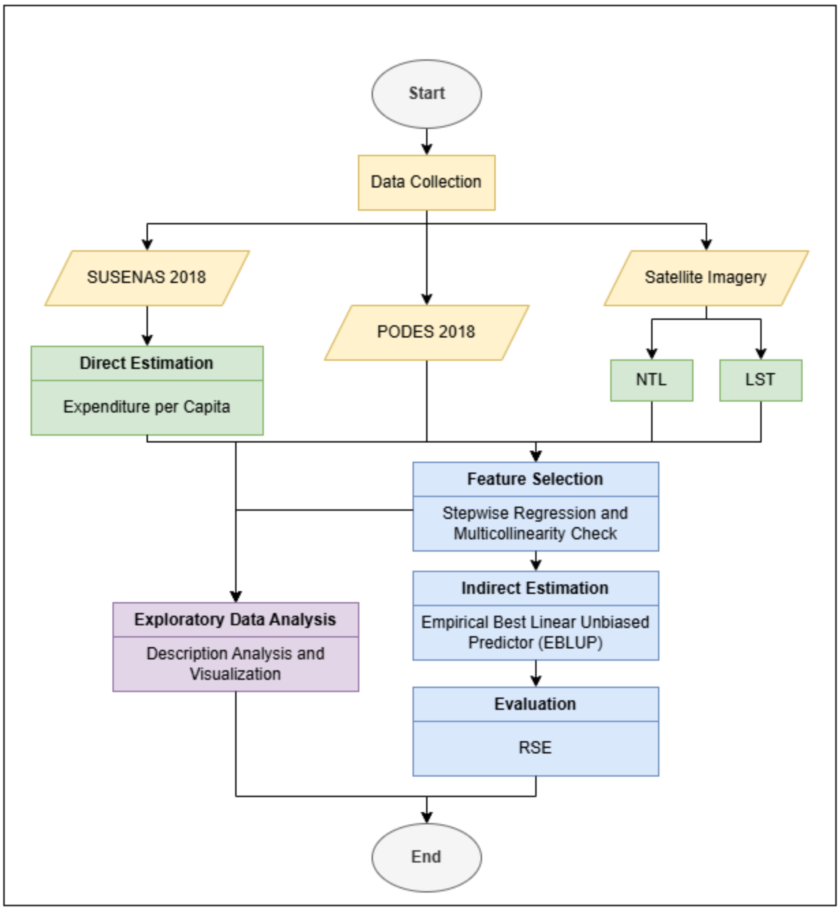


Figure 5. Framework Study

Before the variables are used, the stepwise regression method is used to select which auxiliary variables have a significant effect on the expenditure per capita. Besides stepwise regression, the Pearson correlation is also used to see the relationship between auxiliary variables. The pair of auxiliary variables that have a Pearson correlation value greater than 0.8 is excluded from the model. To ensure that there is no multicollinearity in the auxiliary variables, in addition to the Pearson correlation, the VIF (Variance Inflation Factor) value is also used in feature selection. A good VIF value to use is no more than 10 [30].

The equation of the EBLUP model to estimate expenditure per capita is as follows [5]:

$$\hat{\theta}_i = \boldsymbol{z}_i^T \boldsymbol{\beta} + b_i v_i + e_i \,; i = 1,2,3,\dots,m \tag{2}$$

with random effect area $v_i \sim iid\ N(0, \hat{\sigma}_v^2)$ and error $e_i \sim iid\ N(0, \sigma_e^2)$

$$\hat{\theta}_i^H = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \boldsymbol{z}_i^T \hat{\boldsymbol{\beta}} \tag{3}$$

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2 b_i}{(\hat{\sigma}_v^2 b_i + \psi_i)} \tag{4}$$

where $\hat{\theta}_i^H$ is an estimator of expenditure per capita in i^{th} sub-district using EBLUP, $\hat{\gamma}_i$ is a measure of uncertainly model, $\boldsymbol{z}_i^T$ is a covariate area level that is an auxiliary variable, $\hat{\boldsymbol{\beta}}$ is an estimator of β that is a vector regression coefficient, and $\hat{\sigma}_v^2$ is an estimator of σ_v^2 using the Restricted Maximum Likelihood (REML) method.

Mean Squared Error (MSE) is the average of differences between the true value and the estimator value [5]. The smaller the MSE value, the resulting estimator is considered good and can be used in estimation. According to Rao and Molina [5], MSE from EBLUP model can be obtained by Taylor Series Expansion. By following formula (5), the MSE value for each model can be compared to obtain an MSE value that tends to be low.

$$MSE(\hat{\theta}_i^H) = g_{i1}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{2i}(\hat{\sigma}_v^2) \tag{5}$$

$$g_{1i}(\hat{\sigma}_v^2) = \frac{\hat{\sigma}_v^2 \psi_i}{(\hat{\sigma}_v^2 + \psi_i)} = \hat{\gamma}_i \psi_i \tag{6}$$

$$g_{2i}(\hat{\sigma}_v^2) = (1 - \hat{\gamma}_i)^2 \boldsymbol{z}_i^T \left[\sum_{i=1}^{m} \frac{\boldsymbol{z}_i \boldsymbol{z}_i^T}{(\hat{\sigma}_v^2 + \psi_i)} \right]^{-1} \boldsymbol{z}_i \tag{7}$$

$$g_{3i}(\hat{\sigma}_v^2) = \psi_i^2 (\psi_i + \hat{\sigma}_v^2)^{-3} \underline{V}(\hat{\sigma}_v^2) \tag{8}$$

Where $\underline{V}(\hat{\sigma}_v^2)$ is the asymptotic variance of $\hat{\sigma}_v^2 = 2m^{-2} \sum_{i=1}^{m} (\hat{\sigma}_v^2 + \psi_i)^2$. In addition to Mean Squares Error (MSE), Relative Standard Error (RSE) is utilized for model evaluation; a model with a lower RSE value is more suitable for estimating per capita expenditure. The RSE calculation formula is as follows:

$$RSE\left(\hat{\theta}\right) = \frac{\sqrt{MSE\left(\hat{\theta}\right)}}{\hat{\theta}} \times 100\% \tag{9}$$

where $\hat{\theta}$ is the estimate of expenditure per capita and $MSE(\hat{\theta})$ is the Mean Squared Error (MSE) of the estimator $\hat{\theta}$.

3. Result and Discussion

3.1. Characteristic of Per Capita Expenditure in West Java

Per capita expenditure in West Java is not evenly distributed at the district level. Figure 7 shows that the distribution still contains the high outlier and has a right-skewed. The distribution of level subdistrict expenditures per capita tends to be concentrated in the northwest, with the highest expenditure group (Figure 6). Besides that, the center of West Java is also clustered highest expenditure per capita. This trend is in big cities in West Java, such as Bandung City, Bekasi City, and Depok. The municipalities that exhibited the highest per capita expenditure in West Java in 2018 are encompassed under this group. Sub-districts in these cities tend to cluster homogeneous, conversely, the clustering of sub-districts differs in other cities, including Karawang, Subang, Indramayu, and Cianjur. It is crucial, then, to obtain more precise estimates of per capita expenditures at the subdistrict level.

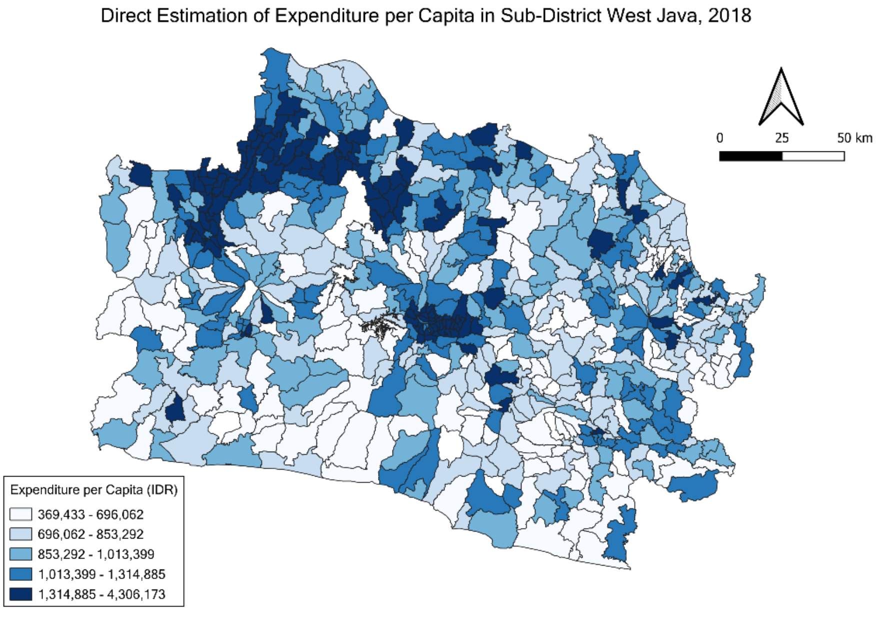


Figure 6. Distribution Expenditure per Capita in Sub-District West Java 2018

Visual mapping is used to find out the general distribution of Nighttime Light Intensity (NTL) and Land Surface Temperature (LST) in West Java. Table 3 shows that the NTL value that has been aggregated in sub-district level is in the index range between 0.24-32.59, whereas the LST value is in the range between 22.62ºC – 38.14ºC. The subdistrict with the lowest NTL value is Naringgul (0.24) in

Cianjur Regency, whereas Sumur Bandung in Bandung City (32.59) has the highest NTL value in West Java. In West Java, the region characterized by the greatest average surface temperature is Bekasi Barat, Bekasi City (38.14°C), whereas Kudadampit, Sukabumi has the lowest temperature (22.62°C).

Nighttime Light Intensity (NTL) is produced by human activity in the night such as building lighting in the town, house lights, gardens, plantation, and limestone quarries [31]. According to Subash et al. [32], NTL demonstrates the social-economy in a given location and can be used to forecast poverty.

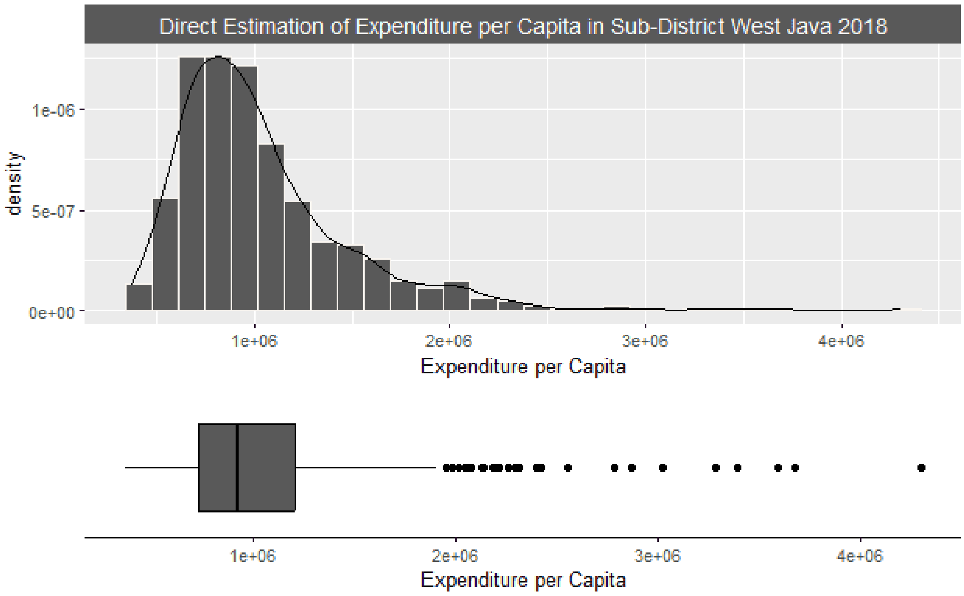


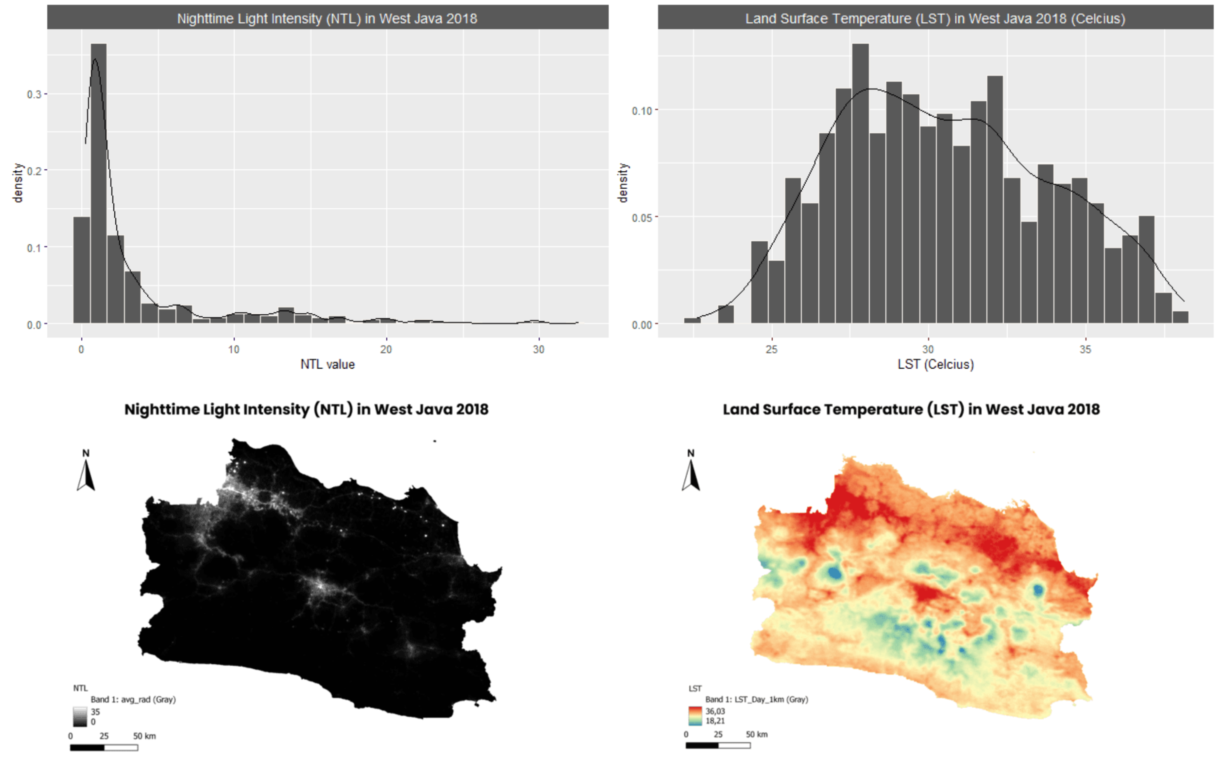
Figure 7. Boxplot and Histogram Expenditure per Capita in Sub-District West Java 2018

Figure 8a shows that the intensity of nighttime light in West Java tends to be high in the northwest and central parts of West Java, Bandung City. This evidence shows that the Nighttime Light Intensity has a high correlation with expenditure per capita. Figure 8b shows that the north part of West Java has a high surface temperature, this is indicated by the red mark in the figure. In contrast with the east side, this area tends to be blue, which indicates that the area has a colder temperature than the north. According to Bodruddoza et al. [33], the higher the Land Surface Temperature (LST) of an area, the more rapid the development and growth of that area will be.

Table 3. Summary of Variables

Variable	Min	Mean	Median	Max	Std Dev	Range
Expenditure per Capita (Y)	369,432.91	1,051,612.26	928,571.89	4,306,173.42	486,660.60	3,936,740.51
University (X1)	0	0	0	7	0.62	7
Islamic Boarding School (X2)	0	50	35	223	41.09	223
Mechanical Training (X3)	0	1	0	48	4.33	48
Language Training (X4)	0	1	0	41	2.67	41
Medical Facility (X5)	0	6	5	27	5.52	27
Mini Market (X6)	0	14	8	190	19.92	190
NTL	0.24	3.72	1.42	32.59	5.43	32.35
LST (Celsius)	22.62	30.49	30.23	38.14	3.31	15.52

Jurnal Aplikasi Statistika & Komputasi Statistik, vol.16(02), pp 205-221 , December, 2024



(a) (b)

Figure 8. (a) Distribution of Nighttime Light Intensity (NTL) in West Java 2018 Sub-district level; (b) Distribution of Land Surface Temperature (LST) in West Java 2018 Sub-district level

3.2. Feature Selection

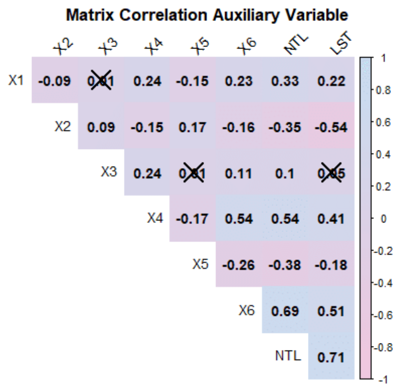


Figure 9. Matrix Correlation Auxiliary Variables

The Pearson correlation between auxiliary variables which is not more than 0.8 (as shown in Figure 9) and the VIF values below 10 for each variable (as shown in Table 4 (2)) indicate that there is no multicollinearity problem. All variables are good predictors because there is no multicollinearity among them, and they have a significant correlation to per capita expenditure.

Table 4. Pearson Correlation and VIF Value

Variable	VIF Value	Pearson Correlation with Y	p-value
X1	1.1323	0.2955	0.0000*
X2	1.5225	-0.3299	0.0000*
X3	1.0788	0.1336	0.0007*
X4	1.6078	0.4520	0.0000*
X5	1.2172	-0.3449	0.0000*
X6	2.0991	0.5432	0.0000*
NTL	3.4628	0.7019	0.0000*
LST	2.6207	0.5648	0.0000*

*significant in 5%

Table 5. Stepwise Regression

Step	AIC
NA	15,610.62
+NTL	15,208.26
+LST	15,199.56
+X5	15,192.03
+X4	15,185.25
+X3	15,181.98
+X1	15,178.88
+X6	15,176.78
+X2	15,174.43

Akaike Information Criterion (AIC) can be used to determine how good a model is. The lower the AIC value, the better the model will be. If all variables are used in this scenario, Table 5 shows that all variables are included in the model, and the minimum AIC is obtained. Therefore, there is no auxiliary variable is dropped in all scenarios.

3.3. Small Area Estimation

The Empirical Best Linear Unbiased Prediction (EBLUP) model is used to estimate the expenditure per capita in West Java, 2018. The scenarios used in this study are: using all PODES variables, using only NTL or LST, using combined remote sensing data, using all variables as auxiliary variables. According to Table 6, every PODES variable is significant at the 5% significant level. This means that all variables from PODES have a significant influence on per capita expenditure in West Java. Similar to Table 6, the next scenarios (shown in Table 7, Table 8, and Table 9) yield significant results in 5%. This implies that the auxiliary variables have done a good job of predicting the per capita expenditure in the four scenarios. In contrast to the previous result, in the scenario where all variables are used in forming the EBLUP model, there is one variable, Language Training (X4) that is not significant at the 5% significant level (Table 10). In SAE modelling, the beta interpretation of each variable is not necessary because the focus of SAE modelling is to improve the estimation accuracy by lowering the Relative Standard Error (RSE) value to as low as possible.

Table 6. EBLUP with PODES as Auxiliary Variables

Variable	Beta	Std Error	t-value	p-value
Intercept	989,954.71	24,506.56	40.40	0.0000*
X1	102,653.68	22,104.02	4.64	0.0000*
X2	-1,887.53	268.85	-7.02	0.0000*
X3	6,254.24	2,669.50	2.34	0.0191*
X4	26,203.50	5,526.00	4.74	0.0000*
X5	-11,027.13	2,081.52	-5.30	0.0000*
X6	7,600.96	690.00	11.02	0.0000*

*significant in 5%

Table 7. EBLUP with NTL as Auxiliary Variable

Variable	Beta	Standard Error	t-Value	p-value
Intercept	780,457.90	12,457.10	62.65	0.0000*
NTL	54,644.50	2,088.44	26.17	0.0000*

*significant in 5%

Table 8. EBLUP with LST as Auxiliary Variable

Variable	Beta	Standard Error	t-Value	p-value
Intercept	-1,105,965.34	120,375.45	-9.19	0.0000*
LST	68512.57	3931.89	17.42	0.0000*

*significant in 5%

Table 9. EBLUP with NTL and LST (Remote Sensing) as Auxiliary Variables

Variable	Beta	Standard Error	t-Value	p-value
Intercept	350,916.79	128,187.91	2.74	0.0062*
NTL	48,070.87	2,845.95	16.89	0.0000*
LST	14,841.07	4,412.08	3.36	0.0008*

*significant in 5%

Table 10. EBLUP with All Variables as Auxiliary Variables

Variable	beta	Standard Error	t-Value	p-value
Intercept	523,226.24	148,806.97	3.52	0.0004*
X1	54,303.92	20,926.54	2.59	0.0095*
X2	-688.22	286.98	-2.40	0.0165*
X3	5,758.84	2,415.87	2.40	0.0166*
X4	9,980.48	5,277.31	1.89	0.0586
X5	-6,234.48	1,962.77	-3.18	0.0015*
X6	2,785.12	751.74	-3.70	0.0002*
NTL	32,851.68	3,689.42	8.90	0.0000*
LST	11,309.70	4,882.52	2.32	0.0205*

*significant in 5%

As shown in Figure 10, certain subdistricts have a high value of direct estimation per capita expenditure. When compared to direct estimation, the outcomes of estimation utilizing the SAE model are comparatively lower. The maximum estimated value obtained using direct estimation is IDR 4,306,173.42, while the maximum estimated value from other models is no more than IDR 2.6 million. The highest value in the direct estimation is in Lengkong Sub-District, Bandung City. In the SAE model, Lengkong is not the sub-district with the highest expenditure per capita.

According to Table 11, there is a tendency for differences in minimum values from direct estimation and indirect estimation models, SAE. The minimum value in direct estimation is IDR 369,433 in Cijati Sub-District, Cianjur. Whereas the minimum value in the indirect estimation with all possibilities dataset is around IDR 390,000. This demonstrates a substantial disparity in the minimum value. Similarly, direct estimating tends to yield a mean figure for expenditure per capita that is greater than that of indirect estimation. At IDR 990,000, the mean value in indirect estimation tends to be uniform. In contrast, direct estimation yields a mean value of IDR 1,051,612.26. The median values tend towards homogeneity. As illustrated in Figure 7's boxplot, the discrepancy between the median and mean values in direct estimation is due to outliers.

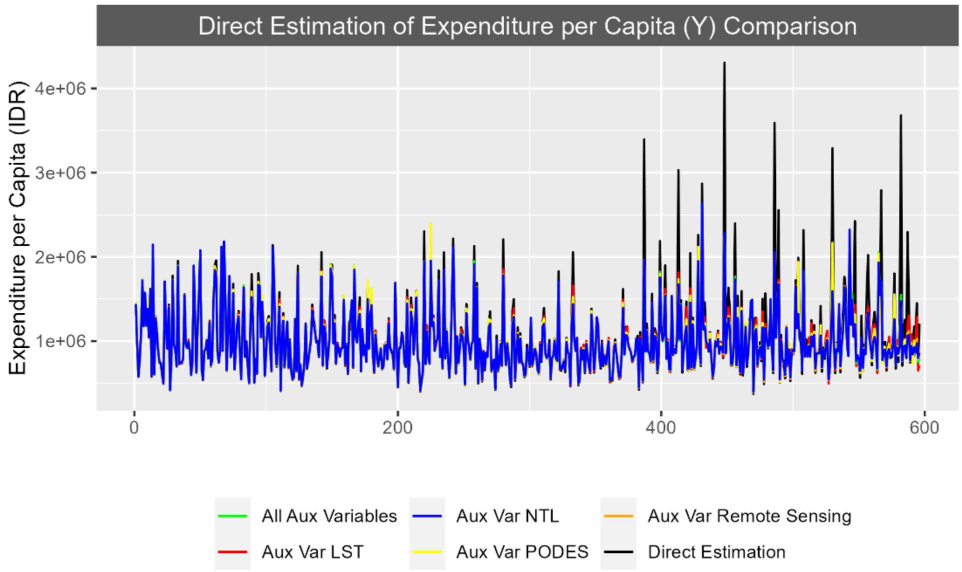


Figure 10. Direct Estimation of Expenditure per Capita Comparison

The SAE model with PODES as the auxiliary variable has a minimum value in Cijati Sub-District (IDR 391,343) Cianjur and a maximum value in Beji Sub-District (IDR 2,400,269). The mean value in this model is IDR 991,298.87. Implementation of remote sensing (NTL) as an auxiliary variable produces the estimation expenditure per capita with the minimum value also in the Cijati Sub-District (IDR 396,827), whereas Sumur Bandung (IDR 2,631,645) in Bandung City is the sub-district with the highest value of expenditure per capita using EBLUP model. The maximum value from the SAE model with Land Surface Temperature (LST) as an auxiliary variable is in East Bekasi Sub-District (IDR 1,977,508), Bekasi City. The maximum estimated value using LST is the minimum value from other models. Meanwhile, the maximum value estimated using a combination of both remote sensing (NTL and LST) and all variables (PODES and remote sensing) is respectively in Sumur Bandung Sub-District. The subdistrict in West Java with the lowest estimated per capita spending is Cijati Sub-District, based on all models employed in the Small Area Estimation.

Table 11. Summary Estimation of Expenditure per Capita in All Model

Indicator	Direct Estimation	Auxiliary Variables Used				
		PODES	**NTL**	LST	Remote Sensing	All Variables
Min	369,432.91	391,343.37	**396,827.26**	389,915.12	397,254.10	397,174.55
Q1	735,262.11	745,966.71	**748,441.52**	742,553.36	747,047.50	747,252.79
Mean	1,051,612.26	991,298.87	**992,055.30**	993,264.59	991,790.79	991,460.83
Median	928,571.89	917,185.09	**906,978.11**	916,912.71	904,545.28	903,064.59
Q3	1,212,535.10	1,162,556.29	**1,137,248.19**	1,178,216.06	1,139,907.94	1,143,204.75
Max	4,306,173.42	2,400,269.65	**2,631,645.32**	1,977,508.37	2,551,022.92	2,341,868.74
Range	3,936,740.51	2,008,926.28	**2,234,818.05**	1,587,593.24	2,153,768.82	1,944,694.19
Std Dev	486,660.60	345,840.83	**356,331.37**	338,007.99	355,409.60	356,292.87

BPS-Statistics Indonesia uses a maximum threshold for the Relative Standard Error (RSE) value that is good to use at 25%. RSE values between 25 and 50 percent are prudent values to use with caution. Furthermore, the resultant estimate lacks reliability since the RSE value exceeds 50%. Figure 11 shows that the RSE from LST model has values that exceed 25%, there are Majalengka (35.17%), Sukahening (30.61%), and Cigugur (28.40%) Sub-Districts. Besides that, Sub-Districts Jatiwaras (25.30%) and Majalengka (25.25%) in PODES as auxiliary variables have RSE values greater than 25%. Majalengka Sub-District also has RSE maximum value in the SAE model with all remote sensing data (25.29%) and all variables dataset (25.03%) as auxiliary variables. The only model that doesn't have an RSE value of more than 25% is the one that uses NTL as an auxiliary variable.

The model containing all variables (10.39%) has the smallest average RSE value among all the proposed models. Followed by the model containing all remote sensing auxiliary variables (10.48%), and then the SAE model with NTL (10.50%) as an auxiliary variable. The minimum RSE value of all models is around 4.7% and the maximum value for all models is around 25%, except for the model with LST as an auxiliary variable (35%).

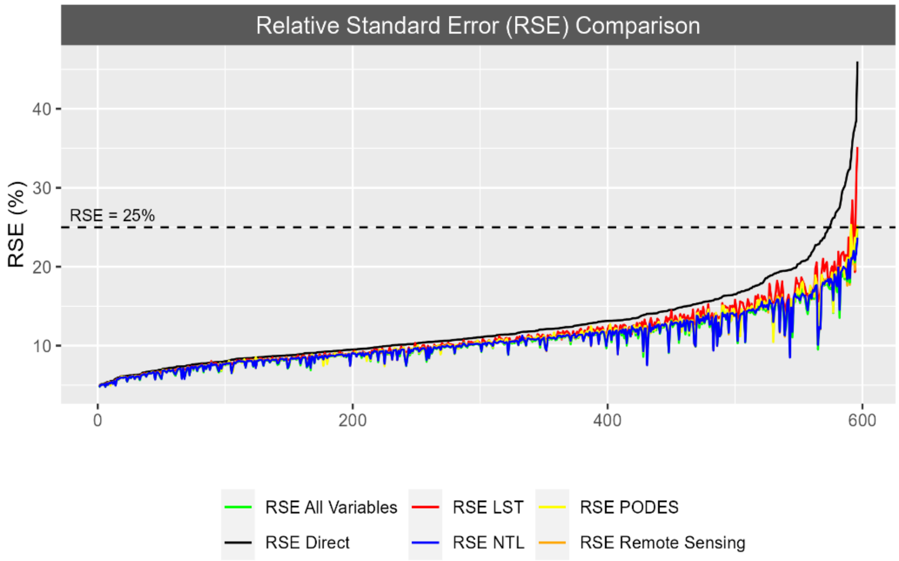


Figure 11. Relative Standard Error (RSE) Comparison

Table 12. Summary of Relative Standard Error (RSE) in All Models (%)

Indicator	Direct Estimation	Auxiliary Variables Used				
		PODES	**NTL**	LST	Remote Sensing	All Variables
Min	5.00	4.77	**4.78**	4.89	4.77	4.71
Q1	8.77	8.23	**8.19**	8.38	8.16	8.12
Mean	12.44	10.70	**10.50**	11.10	10.48	10.39
Median	11.06	10.20	**9.97**	10.41	9.95	9.89
Q3	14.43	12.36	**12.10**	12.88	12.09	11.94
Max	45.99	25.30	**23.68**	35.17	25.29	25.03
Total Sub-District with RSE > 25%	22	2	**0**	3	1	1

Table 13. Normality Test in Error and Random Effect Area

Auxiliary Variables Model EBLUUP	Error	Random Effect Area
PODES	0.0314 (0.4939)	0.0179 (0.9909)
NTL	0.0287 (0.7085)	0.0259 (0.8196)
LST	0.0575 (0.0386)*	0.0339 (0.4992)
Remote Sensing	0.0288 (0.7045)	0.0214 (0.9486)
All Variables	0.0519 (0.0804)	0.0257 (0.8244)

* significant in 5%

In formula (2), the errors ($e_i \sim iid\ N(0, \sigma_e^2)$) and the random effect area ($v_i \sim iid\ N(0, \sigma_v^2)$ of EBLUP are distributed in a normal distribution [5]. Kolmogorov Smirnov is a normality test by comparing one distribution with another distribution, in this case the distribution of error (e_i) and random effect area (v_i) are compared with the normal distribution. The null hypothesis of the normality test is H_0: Error (e_i) or random effect area (v_i) is not normally distributed and the alternative hypothesis is H_a: Error (e_i) or random effect area (v_i) is normally distributed.

Table 13 shows that only the EBLUP model with LST as an auxiliary variable has errors (e_i) that are not spread in a normal distribution. Whereas all random effect areas are normally distributed.

Table 14. Model Evaluation

Auxiliary Variables Model EBLUP	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)
PODES	16,721	16,747
NTL	**16,637**	**16,651**
LST	16,842	16,855
Remote Sensing	16,627	16,645
All Variables	16,591	16,635

Based on Table 14, the model with all variables used has the lowest value of AIC and BIC. This means that the estimation with all variables is better than the others. However, modelling with all variables does not fulfill the principle of parsimony. The model using NTL and all remote sensing variables as auxiliary variable has better AIC and BIC values than the PODES model. Hence, this provides an opportunity that the remote sensing data can be used as auxiliary variables in the estimation expenditure per capita using EBLUP.

Based on the description above, the estimation of expenditure per capita using the EBLUP model with Nighttime Light Intensity (NTL) is the best model. The estimated value using EBLUP with NTL tends to have similar characteristics and represent to the true value. The RSE values is relatively lower than direct estimation and especially in this model all RSE values produced has been in category that can be used for the government.

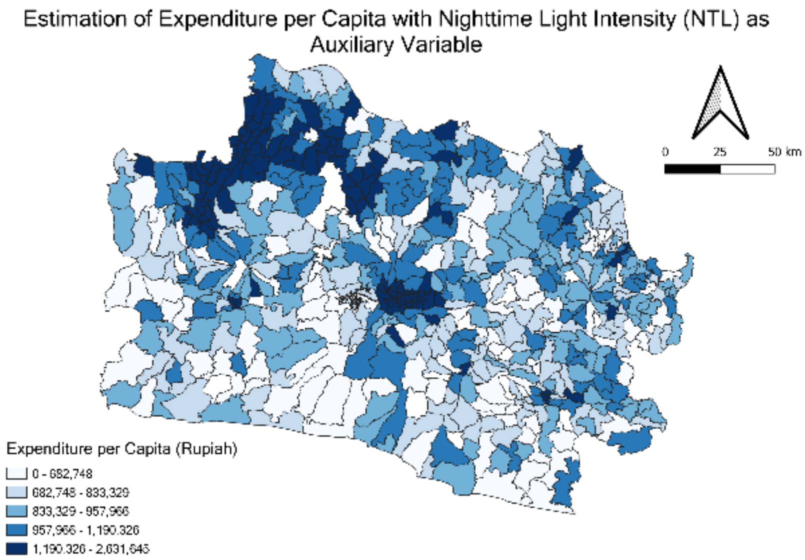


Figure 12. Estimation Expenditure per Capita Best EBLUP model with NTL as auxiliary variables

Figure 12 shows the results best model of estimation expenditure per capita, Empirical Best Linear Unbiased Prediction (EBLUP) with Nighttime Light Intensity (NTL) as an auxiliary variable. The Norwest and the central regions of West Java province have relatively high expenditure per capita. On the other hand, the Southwest region tends to have a low value of expenditure per capita. Therefore, based on the above analysis, estimation using Empirical Best Linear Unbiased Prediction (EBLUP) with Nighttime Light Intensity (NTL) tends to be similar to direct estimation and has the advantage of relatively lower RSE values than direct estimation.

4. Conclusion

The problem of the high Relative Standard Error (RSE) value per capita expenditure at the sub-district level in West Java can be overcome by using the Small Area Estimation (SAE) model. The proposed model can reduce the RSE value of the estimate. The best model obtained is the Empirical Best Linear Unbiased Prediction (EBLUP) model using Nighttime Light Intensity (NTL) as an auxiliary variable. Therefore, there is potential for remote sensing data to be utilized as an auxiliary in the SAE model. This remote sensing data has the advantage of data that is relatively quick in updating and cheap compared to conventional census data.

The government may consider that remote sensing data can be utilized as auxiliary variables in estimating expenditure per capita in particularly and estimating other economics indicators in general using the Small Area Estimation (SAE) model. Other studies may consider exploring other remote sensing data in developing auxiliary variables in the Small Area Estimation (SAE) model.

Ethics approval

Not required.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable feedback.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Derived data supporting the findings of this study are available from the corresponding author on request.

Credit Authorship

Muhamad Feriyanto: Conceptualization, Data Collection, Formal Analysis, Writing - Original Draft, Visualization**.** **Arie Wahyu Wijayanto**: Methodology, Writing - Review & Editing, Supervision**.** **Ika Yuni Wulansari**: Writing - Review & Editing, Supervision. **Novia Budi Parwanto**: Writing - Review & Editing, Supervision.

References

[1] (BPS) Badan Pusat Statistik, “Konsumsi dan Pengeluaran,” Konsumsi dan Pengeluaran [Consumption and Expenditure]. Accessed: Nov. 30, 2023. [Online]. Available: https://www.bps.go.id/subject/5/konsumsi-dan-pengeluaran.html

[2] (BPS) Badan Pusat Statistik, “Ekonomi Indonesia 2018 Tumbuh 5,17 Persen,” Ekonomi Indonesia 2018 Tumbuh 5,17 Persen [Indonesia's 2018 Economy grew 5.17 Percent]. Accessed: Nov. 30, 2023. [Online]. Available: https://www.bps.go.id/pressrelease/2019/02/06/1619/ekonomi-indonesia-2018-tumbuh-5-17-persen.html

[3] (BPS) Badan Pusat Statistik, “Statistik Indonesia 2018,” Statistik Indonesia 2018 [Statistics Indonesia 2018]. Accessed: Nov. 30, 2023. [Online]. Available: https://www.bps.go.id/publication/2018/07/03/5a963c1ea9b0fed6497d0845/statistik-indonesia-2018.html

[4] Badan Pusat Statistik, *PEDOMAN KEPALA BPS PROVINSI, STATISTISI AHLI MADYA/KOORDINATOR FUNGSI STATISTIK SOSIAL BPS PROVINSI, DAN KEPALA BPS KABUPATEN/KOTA SUSENAS MARET 2022* [*GUIDELINES FOR THE HEAD OF THE*

PROVINCIAL BPS, INTERMEDIATE EXPERT STATISTICIAN/COORDINATOR OF THE SOCIAL STATISTICS FUNCTION OF THE PROVINCIAL BPS, AND THE HEAD OF THE DISTRICT / CITY BPS SUSENAS MARCH 2022]. Badan Pusat Statistik, 2022.
[5] J. N. K. Rao and I. Molina, *Small area estimation*, Second edition. in Wiley series in survey methodology. Hoboken, New Jersey: John Wiley & Sons, Inc, 2015.
[6] D. T. Lindgren, *Land use planning and remote sensing*. in Remote sensing of earth resources and environment, no. 2. Dordrecht ; Boston : Hingham, MA, USA: M. Nijhoff Publishers ; Distributors for the U.S. and Canada, Kluwer Academic Pub, 1985.
[7] S. Hutasavi and D. Chen, "Estimating District-Level Electricity Consumption Using Remotely Sensed Data in Eastern Economic Corridor, Thailand," *Remote Sensing*, vol. 13, no. 22, p. 4654, Nov. 2021, doi: 10.3390/rs13224654.
[8] S. R. Putri, A. W. Wijayanto, and A. D. Sakti, "Developing Relative Spatial Poverty Index Using Integrated Remote Sensing and Geospatial Big Data Approach: A Case Study of East Java, Indonesia," *IJGI*, vol. 11, no. 5, p. 275, Apr. 2022, doi: 10.3390/ijgi11050275.
[9] S. Shi, Y. Ye, and R. Xiao, "Evaluation of Food Security Based on Remote Sensing Data—Taking Egypt as an Example," *Remote Sensing*, vol. 14, no. 12, p. 2876, Jun. 2022, doi: 10.3390/rs14122876.
[10] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping," *AAAI*, vol. 30, no. 1, Mar. 2016, doi: 10.1609/aaai.v30i1.9906.
[11] J. Yin, Y. Qiu, and B. Zhang, "Identification of Poverty Areas by Remote Sensing and Machine Learning: A Case Study in Guizhou, Southwest China," *IJGI*, vol. 10, no. 1, p. 11, Dec. 2020, doi: 10.3390/ijgi10010011.
[12] S. R. Putri, A. W. Wijayanto, and S. Pramana, "Multi-source satellite imagery and point of interest data for poverty mapping in East Java, Indonesia: Machine learning and deep learning approaches," *Remote Sensing Applications: Society and Environment*, vol. 29, p. 100889, Jan. 2023, doi: 10.1016/j.rsase.2022.100889.
[13] S. Ahmed, "Assessment of urban heat islands and impact of climate change on socioeconomic over Suez Governorate using remote sensing and GIS techniques," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 21, no. 1, pp. 15–25, Apr. 2018, doi: 10.1016/j.ejrs.2017.08.001.
[14] Y. Zheng, Q. Zhou, Y. He, C. Wang, X. Wang, and H. Wang, "An Optimized Approach for Extracting Urban Land Based on Log-Transformed DMSP-OLS Nighttime Light, NDVI, and NDWI," *Remote Sensing*, vol. 13, no. 4, p. 766, Feb. 2021, doi: 10.3390/rs13040766.
[15] Y. Sun, S. Wang, X. Zhang, T. O. Chan, and W. Wu, "Estimating local-scale domestic electricity energy consumption using demographic, nighttime light imagery and Twitter data," *Energy*, vol. 226, p. 120351, Jul. 2021, doi: 10.1016/j.energy.2021.120351.
[16] L. Lin, L. Di, C. Zhang, L. Guo, and Y. Di, "Remote Sensing of Urban Poverty and Gentrification," *Remote Sensing*, vol. 13, no. 20, p. 4022, Oct. 2021, doi: 10.3390/rs13204022.
[17] G. E. Battese, R. M. Harter, and W. A. Fuller, "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data," *Journal of the American Statistical Association*, vol. 83, no. 401, pp. 28–36, Mar. 1988, doi: 10.1080/01621459.1988.10478561.
[18] R. Singh, D. P. Semwal, A. Rai, and R. S. Chhikara, "Small area estimation of crop yield using remote sensing satellite data," *International Journal of Remote Sensing*, vol. 23, no. 1, pp. 49–56, Jan. 2002, doi: 10.1080/01431160010014756.
[19] G. F. Santoso and S. Muchlisoh, "Estimasi Produktivitas Padi Level Kecamatan di Kabupaten Tulungagung Menggunakan Geoadditive SAE" [Sub-district Level Rice Productivity Estimation in Tulungagung Regency Using Geoadditive SAE], *asks*, vol. 12, no. 3, pp. 23–36, Mar. 2022, doi: 10.34123/jurnalasks.v14i1.385.
[20] N. T. Longford, M. G. Pittau, R. Zelli, and R. Massari, "Poverty and inequality in European regions," *Journal of Applied Statistics*, vol. 39, no. 7, pp. 1557–1576, Jul. 2012, doi: 10.1080/02664763.2012.661705.
[21] T. Siswantining, "GEOINFORMATIKA PADA KASUS AREA KECIL DAN PENERAPANNYA UNTUK MENDETEKSI KANTONG- KANTONG KEMISKINAN DI JEMBER" [GEOINFORMATICS IN THE CASE OF SMALL AREAS AND ITS

APPLICATION TO DETECT POCKETS OF POVERTY IN JEMBER], *Institute Pertanian Bogor Pasca Sarjana Statistika*, 2013.

[22] N. Istiana, “SMALL AREA ESTIMATION WITH EXCESS ZERO (CASE STUDY: INFANT MORTALITY RATE IN JAVA ISLAND),” *asks*, vol. 13, no. 1, pp. 25–34, Sep. 2021, doi: 10.34123/jurnalasks.v13i1.270.

[23] S. Guha and H. Chandra, “Measuring disaggregate level food insecurity via multivariate small area modelling: evidence from rural districts of Uttar Pradesh, India,” *Food Sec.*, vol. 13, no. 3, pp. 597–615, Jun. 2021, doi: 10.1007/s12571-021-01143-1.

[24] Md. J. Hossain, S. Das, H. Chandra, and M. A. Islam, “Disaggregate level estimates and spatial mapping of food insecurity in Bangladesh by linking survey and census data,” *PLoS ONE*, vol. 15, no. 4, p. e0230906, Apr. 2020, doi: 10.1371/journal.pone.0230906.

[25] P. A. Kaban, B. I. Nasution, R. E. Caraka, and R. Kurniawan, “Implementing night light data as auxiliary variable of small area estimation,” *Communications in Statistics - Theory and Methods*, pp. 1–18, May 2022, doi: 10.1080/03610926.2022.2077963.

[26] A. Ubaidillah, K. A. Notodiputro, A. Kurnia, and I. W. Mangku, “Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia,” *Journal of Applied Statistics*, vol. 46, no. 15, pp. 2845–2861, Nov. 2019, doi: 10.1080/02664763.2019.1615420.

[27] A. Asian Development Bank, “How Nighttime Lights Help Us Study Development Indicators | Asian Development Blog,” How Nighttime Lights Help Us Study Development Indicators. Accessed: Nov. 04, 2024. [Online]. Available: https://blogs.adb.org/blog/how-nighttime-lights-help-us-study-development-indicators

[28] R. Beyer, J. Yao, and Y. Hu, “Measuring Quarterly Economic Growth from Outer Space,” *IMF Working Papers*, vol. 2022, no. 109, p. 1, Jun. 2022, doi: 10.5089/9798400211553.001.

[29] S. Yu *et al.*, “Land Surface Temperature Changes in Different Urbanization Increments in China since 2000,” *Land*, vol. 13, no. 4, p. 417, Mar. 2024, doi: 10.3390/land13040417.

[30] M. H. Kutner, Ed., *Applied linear statistical models*, 5th ed. in The McGraw-Hill/Irwin series operations and decision sciences. Boston: McGraw-Hill Irwin, 2005.

[31] Q. Zhang and K. C. Seto, “Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data,” *Remote Sensing of Environment*, vol. 115, no. 9, pp. 2320–2329, Sep. 2011, doi: 10.1016/j.rse.2011.04.032.

[32] S. P. Subash, R. R. Kumar, and K. S. Aditya, “Satellite data and machine learning tools for predicting poverty in rural India,” *Agri. Econ. Rese. Revi.*, vol. 31, no. 2, p. 231, 2018, doi: 10.5958/0974-0279.2018.00040.X.

[33] Md. Bodruddoza Mia, Rahul Bhattacharya, and ASM Woobaidullah, “Correlation and Monitoring of Land Surface Temperature, Urban Heat Island with Land use-land cover of Dhaka City using Satellite imageries,” *International Journal of Research in Geography*, vol. 3, no. 4, 2017, doi: 10.20431/2454-8685.0304002.