Poverty Mapping of the Coastal Areas Using Spatial Empirical Best Linear Unbiased Prediction Method

E Sunandi, D Agustina and H Fransiska
Universitas Bengkulu, Jl. W.R. Supratman, Bengkulu 38125, Indonesia
E-mail: esunandi@unib.ac.id

Abstract. This research aims to map out the poverty data in the Coastal Areas of Mukomuko District. The estimation method used is Spatial Empirical Best Linear Unbiased Prediction (EBLUP). So, the poverty estimator in the coastal area of Mukomuko District is obtained. The evaluation of parameter estimator is calculated by the value of MSE (Mean Square Error) using Bootstrap resampling method. Based on the result of the study is seen that the MSE value of Spatial EBLUP estimators is smaller than the MSE value of the direct estimator (Y) in each village. Finally, the estimation is presented as poverty map. The estimation of spatial correlation is positive and strong (0.52).

1. Introduction
Poverty is a problem that is faced by the Indonesian people to date. According to Bereau of Statistic (BPS) from the Bengkulu Central Statistics Bereau, in March 2016, the number of poor people in Bengkulu Province reached 328.61 thousand people (17.32%). The percentage of poor people in urban areas has decreased. However, in the Village, there was an increase. The percentage of poor people in urban areas in March 2016 was 16.19% and 17.85% in the rural areas. Some rural areas in Bengkulu province are in the coastal areas. Sadly, the level of poverty in the coastal communities in Indonesia is still very worrying by 32.4% [1].

Fishermen’s poverty problem is a multidimensional problem so government needs a complete solution to solve it. There are several aspects that caused poverty of fishermen or coastal communities, including: Government policies that do not take a side to the poor, many policies related to the poverty reduction are top down and always make the community as an object, not a subject. Another condition, which deteriorated the level of fishermen prosperous, is the lack of poverty data. As a result, the statistics data are neglected and inaccurate.

The statistical method is used as an effort to get accuracy about poverty information is Small Area Estimation (SAE). The SAE method is a statistical method for estimating parameters in a subpopulation where the number of samples is small or nonexistent. This method utilizes data from large domains to predict the variables that concern the smaller domain.

The models in SAE assume that the random effects of area errors are mutually independent. But in some cases, this assumption is often violated [2]. The main reason is the diversity of an area affected by the surrounding area so that spatial effects can be incorporated into random influences. Spatial effects are common between one area and another, which means that one area affects the other. In statistics, a model explaining the relationship between an area and the surrounding area is a spatial model.
SAE’s researches using spatial effects have been carried out, see [2-4]. Spatial effect could improve parameter estimation in a small area indicated by a decrease of the RMSE value (Root Mean Square Error) [2]. A study was conducted by using the best spatial linear bias prediction method for estimating outcome per capita [3]. Another researcher conducts a study on logarithmic transformation in the Spatial EBLUP method [4].

Till now, the SAE research using Spatial EBLUP methods in coastal areas was rarely found. Therefore, the authors are really interested in mapping the level of poverty in the coastal area of Mukomuko District by using the Spatial Empirical Best Linear Unbiased Prediction Method. This method is used because of the independence assumption violated. By using spatial effects, the resulting estimation will be better. For models with assume $e_\sim$Normal, this model is the best model.

2. Materials and Methods

This study uses data from the Agency for the Implementation of Food Counseling and Food Security, and the Office of Marine and Fisheries of the Mukomuko District. BPS data used is Podes Mukomuko District in 2014. In addition, the determination of villages which are coastal areas also uses BPS data "Districts in Figures". Mukomuko District consists of 152 villages. 13.16% or 20 villages of all are the villages which located in the coastal areas.

The research variables used are the percentage of poor families (Y), Number of malnutrition Sufferers in the Last 3 Years (X1), Number of Deaths of Mothers When Childbirth (X2), Number of Families living in slums (X3), Number of Health Facilities (X4), Number of Recipients of Public Health Insurance (Jamkesmas) (X5), Number of Not Affordable Certificates (SKTM) (X6), Number of BPJS Receipts (X7), Number of Family without Electricity (X8), and Number of disabilities (X9).

2.1. Poverty

The poor are people who have an average per capita expenditure per month below the poverty line (GK). Technically, GK is built from two components, namely the Food Poverty Line (GKM) and the Non-Food Poverty Line (GKNM). GKM is the expenditure value of minimum food requirements which is equal to 2,100 kilo calories per capita per day; meanwhile, GKNM is a minimum requirement for housing, clothing, education and health [5].

In addition, BPS defines household outcome a month as all household expenses for a month to meet consumption needs for all household members. Outcome Per capita data is obtained from the total monthly household outcome divided by the number of household members [5]. Based on the assumption, the sampling is based on drawing a simple random sample, the average of outcome per capita of a village is obtained by the formula:

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

where:

$\bar{y}_i$ = the average of outcome village per capita, $i = 1, 2, ..., m$

$y_{ij}$ = outcome of the household-j per capita in village to-i and to-j = 1, 2, ..., $n_i$

$n_i$ = number of households in the village to-i

$m$ = number of villages

2.2. Small Area Estimation (SAE)

Small area estimation or SAE is a statistical technique to estimate the parameters of subpopulations with small sample sizes [6]. In Indonesia, these subpopulations can be in the form of provinces, districts/cities, sub-districts or villages. In general, there are three approaches to get SAE parameter estimators, namely direct estimation, indirect estimation, and composite estimation.

SAE technique is an indirect estimation that combines survey data with other supporting data, for example from previous census data that contains variables with the same characteristics as survey data so that they can be used to estimate smaller areas and provide a better level of accuracy. The indirect
estimation process is estimation in a domain by connecting information in that area with other areas through a right model. This means that these allegations include data from other domains.

In SAE, there are used two types of basic models, namely area-based models and unit-based models [6]. In the area-based small area estimation model, the supporting data available is only up to the area level. The area level model connects the estimator directly to a small area with supporting data from another domain for each area.

The small area parameter that to observe is \( \theta_i \), index \( i \) shows area index. The linear model that explains the relationship is:

\[
\theta_i = x_i^T \beta + z_i v_i
\]  \hspace{1cm} (1)

where:

- \( \beta = (\beta_1, ..., \beta_p)^T \) is the regression coefficient \( p \times l \)
- \( z_i \) = known positive constants to-\( i \)
- \( v_i \) = random area effect, assumed \( v_i \sim iid N(0, \sigma^2) \), \( i = 1, 2, ..., m \)

In making conclusions about the population, it is assumed that the estimated value is direct \( \hat{\theta}_i \) is known then it can be stated as follows:

\[
\hat{\theta}_i = \theta_i + e_i
\]  \hspace{1cm} (2)

Random variable \( e_i \) is sampling error, assumed \( e_i \sim iid N(0, \psi_i) \) and \( i = 1, 2, ..., m \).

A small area estimation model for the area level consists of two model component levels, namely the indirect estimation model component according to equation (1) and the direct estimation model component according to equation (2). The models in equations (1) and (2) if combined form the following equation:

\[
\hat{\theta}_i = x_i^T \beta + z_i v_i + e_i
\]  \hspace{1cm} (3)

2.3. Spatial Empirical Best Linear Unbiased Predictor (SEBLUP) Method

One method used in building the SAE model is based on the mixed linear model. The SEBLUP method is a parameter estimation that minimizes the Mean Square Error (MSE) among classes of estimation of other non-linear linear parameters by adding spatial influence. SEBLUP estimator \((\tilde{\sigma}_u^2, \tilde{\rho})\) is obtained from Spatial BLUP by changing the value \(\sigma_u^2, \rho\) with the estimator. Normal assumptions from random influences are used to estimate \(\sigma_u^2\) and \(\rho\) by using both ML and REML procedures with the log-likelihood function having a global maximum and some local maximum [7][8].

The estimation can be obtained iteratively by using the Nelder-Mead algorithm [9] and scoring algorithms. The application of these two procedures sequentially needs to be done because the Nelder-Mead algorithm method does not depend on the selection of the starting point but it is not too efficient and the results obtained approach the global maximum, while the scoring algorithm requires the right starting point to get the maximum function. The results of the estimation are then used to estimate SEBLUP as follows:

\[
\tilde{\theta}_i^*(\tilde{\sigma}_u^2, \tilde{\rho}) = x_i \tilde{\beta} + b_i^T \{\tilde{\sigma}_u^2[(I - \tilde{\rho}W)(I - \tilde{\rho}W^T)]^{-1}\}Z^T
\]

\[
\times \{\text{diag}(\tilde{\sigma}_u^2) + Z \tilde{\sigma}_u^2[(I - \tilde{\rho}W)(1 - \tilde{\rho}W^T)]^{1}\}^{-1}(\tilde{\theta} - X \tilde{\beta})
\]

\[
\text{MSE}[\tilde{\theta}_i^*(\tilde{\sigma}_u^2, \tilde{\rho})] \text{ for the SEBLUP model with a random effect normal distribution is:}
\]

\[
\text{MSE}[\tilde{\theta}_i^*(\tilde{\sigma}_u^2, \tilde{\rho})] = \text{MSE}[\tilde{\theta}_i^*(\tilde{\sigma}_u^2, \rho)] + E[\tilde{\theta}_i^*(\tilde{\sigma}_u^2, \tilde{\rho}) - \tilde{\theta}_i^*(\tilde{\sigma}_u^2, \rho)]^2
\]

The form \(E[\tilde{\theta}_i^*(\tilde{\sigma}_u^2, \tilde{\rho}) - \tilde{\theta}_i^*(\tilde{\sigma}_u^2, \rho)]^2\) is estimated by Taylor and symbolized by \(g_{3i}(\tilde{\sigma}_u^2, \rho)\) (Kackar and Harville 1984 referenced in [8]), that is:
The results of the estimation of the percentage of poor families can be seen in Table 1.

Table 1. Comparison of Direct Estimator and SEBLUP Estimator Statistics

|                | $\hat{\theta}^{DE}$ | $\hat{\theta}^{ML}$ | $\hat{\theta}^{REML}$ |
|----------------|----------------------|----------------------|------------------------|
| Mean           | 8.56                 | 7.92                 | 8.20                   |
| Median         | 8.24                 | 7.63                 | 7.84                   |
| Standard Deviation | 8.71              | 8.37                 | 8.49                   |
| Sample Variance | 75.89              | 69.98                | 72.07                  |
| Minimum        | 0.00                 | 0.00                 | 0.00                   |
| Maximum        | 35.36                | 35.43                | 35.47                  |

Comparison of observation data and estimation results of the percentage of poor families through SEBLUP in each village in Mukomuko District can be seen in Figure 2. From Figure 2 it can be seen that the estimation results using the small area estimation model with the SEBLUP method has the same trend with the data observation. The obtained model has good flexibility, can be seen from the plot of the alleged results that are able to follow the pattern of distribution of observation data.
Figure 1. Map of the Distribution of Coastal Areas in Mukomuko District

Table 2. Comparison of MSE of several Estimation Methods

|                  | $\hat{\theta}^{ML}$ | $\hat{\theta}^{REML}$ | $\hat{\theta}^{DE}$ |
|------------------|----------------------|------------------------|---------------------|
| Mean             | 7.41                 | 6.63                   | 7.17                |
| Standard Error   | 1.11                 | 0.99                   | 1.05                |
| Median           | 8.07                 | 7.54                   | 8.69                |
| Standard Deviation | 4.97               | 4.42                   | 4.71                |
| Sample Variance  | 24.73               | 19.51                  | 22.18               |
| Minimum          | 0.00                 | 0.00                   | 0.00                |
| Maximum          | 17.60               | 15.95                  | 15.94               |
From Table 2, it can be seen that the MSE value of the SEBLUP ML estimator and SEBLUP REML is smaller than the MSE value from the direct estimator \((Y)\) for each village. This indicates that the estimation with the SEBLUP method can improve the estimation of parameters obtained using the direct method.

In Table 3, we can see the regression coefficient estimator value with the SEBLUP ML and REML methods. Estimation values for the resulting spatial coefficients are positive and strong \((0.52)\), meaning that a village in Mukomuko District which has a large percentage of poor families which is surrounded by other villages that have a large percentage of poor families, and a village has the percentage of poor families is surrounded by other villages that have a small percentage of poor families too. This is also supported by the small area variance random effect estimator values. Estimation of random effect variants through ML is smaller than REML \((3.45)\).
Table 3. Comparison of parameter estimators

|                      | $\hat{\theta}_i^{ML}$ |                      | $\hat{\theta}_i^{REML}$ |
|----------------------|------------------------|----------------------|--------------------------|
|                      | beta       | std.error | pvalue | beta       | std.error | pvalue |
| (Intercept)          | 10.26      | 3.59      | 0.00   | 10.58      | 4.60      | 0.02   |
| X1                   | 1.18       | 0.61      | 0.05   | 1.11       | 0.77      | 0.15   |
| X2                   | -0.65      | 1.54      | 0.67   | -0.29      | 1.90      | 0.88   |
| X3                   | 0.38       | 0.31      | 0.21   | 0.38       | 0.38      | 0.32   |
| X4                   | -0.92      | 0.41      | 0.03   | -0.93      | 0.53      | 0.08   |
| X5                   | -0.01      | 0.01      | 0.44   | -0.01      | 0.02      | 0.54   |
| X6                   | -0.01      | 0.02      | 0.66   | -0.01      | 0.03      | 0.64   |
| X7                   | 0.00       | 0.01      | 0.59   | 0.00       | 0.01      | 0.70   |
| X8                   | 0.27       | 0.05      | 0.00   | 0.26       | 0.06      | 0.00   |
| X9                   | -0.06      | 0.16      | 0.70   | -0.05      | 0.20      | 0.82   |
| Spatial Correlation  | 0.52       |           |        | 052        |           |        |
| Variant of random    | 3.45       |           |        | 7.32       |           |        |
| influences           |            |           |        |            |           |        |

Figure 3. Poverty Mapping of the Coastal Areas of Mukomuko District
Poverty mapping of the coastal areas of Mukomuko District can be seen in the picture above. On this map, the poverty level is divided based on the data quintile. Then, given the coloring on the sub-district polygon with the color gradation according to the poverty level. Districts that have a high percentage of poor families, have the brightest colored polygon. Whereas sub-districts which have a low percentage of poor families, have dark purple polygons.

It can be seen that the spread of poverty in Mukomuko District spread randomly. The highest poverty rates are in Air Rami and Ipuh Subdistricts. Whereas sub-districts which have 75% of poor families are Penarik Subdistrict and XIV Koto District. Teramang Jaya sub-district has 50% of poor families. Finally, sub-districts that have low poverty rates (≤ 25%) are Kecamatan Mukomuko and Air Dikit.

4. Conclusion
The estimation results using a small area estimation model with the SEBLUP method have the same trend (trend) as data. The model obtained has good flexibility, can be seen from the predicted plot that is related to the distribution pattern of the data. In addition, the MSE value of the SEBLUP ML and SEBLUP REML estimators is smaller than the MSE value of the direct estimator (Y) for each the village. This between estimation with the SEBLUP method can improve the estimation of parameters obtained using the direct method.

In addition, the estimation value for the resulting spatial coefficient is positive and strong (0.52), meaning that a village in Mukomuko District has a large percentage of poor families, surrounded by other villages that have a large percentage of poor families, and a village who have a small percentage of poor families, surrounded by other villages that have a small percentage of poor families. This is also supported by a small area random effect variant estimator value. Estimation of random effect variants through ML is smaller than REML (3.45).

The spread of poverty levels in Mukomuko District spread randomly. The highest poverty level is in Air Rami and Ipuh Subdistricts. Whereas sub-districts that have 75% of poor families are Penarik District and Koto District XIV. Teramang Jaya District has 50% of poor families. Finally, sub-districts that have low poverty rates (≤ 25%) are Kota Mukomuko District and Air Dikit District.

5. Acknowledgment
This research is supported by Universitas Bengkulu under grant Penelitian Pembinaan Universitas Bengkulu 2018 number 1538/UN30.15/LT/2018.

References
[1] Kristiyanti M 2016 Pemberdayaan Masyarakat Pesisir Pantai Melalui Pendekatan ICZM (Integrated Coastal Zone Management) Proceeding: Seminar Nasional Multi Disiplin Ilmu Unisbank (Sendi_U) Ke-2 ISBN : 978-979-3649-96-2
[2] Sunandi E 2014 Model Logit Normal dengan Efek Spasial pada Pendugaan Area Kecil Prosiding Semirata 2014 Bidang MIPA BKS Barat ISBN: 978-602-70491-0-9 p 98
[3] Matualage D, Saefuddin A, Wigena AH 2011 Pendekatan Small Area Estimation untuk Menduga Pengeluaran Per Kapita Rumah Tangga Tiap Desa dengan Empirical Best Linear Unbiased Prediction (Studi kasus: Kabupaten Jember Provinsi Jawa Timur) Prosiding Seminar Nasional Statistika Universitas Diponegoro ISBN: 978-979-097-142-4 p 655
[4] Zainuddin HA, Notodiputro KA, Sadik K.A 2015 Study of Logarithmic Transformation Model in Spatial Empirical Best Linear Unbiased Prediction (SEBLUP) Method of Small Area Estimation Proceedings of ”The 7Th SEAMS-UGM Conference 2015” p 67
[5] BPS 2015 Indicator Kesejahteraan Rakyat Provinsi Bengkulu (Bengkulu: BPS)
[6] Rao JNK 2003 *Small Area Estimation* (London: Wiley)

[7] Petrucci A, Salvati N 2004 Small Area Estimation Using Spatial Information. The Rathbun Lake Watershed Case Study *Dipartimento di Statistica”G. Parenti” vialemorgagni* 59-50134

[8] Pratesi M, Salvati N 2008 Small Area Estimation: the EBLUP Estimator Based On Spatially Correlated Random Area Effects *Statistical methods and applications, Stat. Meth.& Appl* 17 p113

[9] Nelder, John A, R Mead 1965 A Simplex Method for Function Minimization *Computer Journal* 7 p 308