Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 5 (2016), pp. 3977-3989
© Research India Publications
http://www.ripublication.com/gjpam.htm

Optimum Spatial Weighted in Small Area Estimation

Asfar, Anang Kurnia* and Kusman Sadik

Departement of Statistics
Bogor Agricultural University
Jalan Pajajaran, Kampus IPB Baranangsiang, Bogor 16128, Indonesia.
**Corresponding author:*

Abstract

Spatial weighted matrix is an important component in spatial data modeling. Selection of spatial weighting matrix will provide a sensitive analysis results in many different types of studies. In addition, determining weighting matrix is an important and fundamental thing in obtaining an accurate estimation. Referring to this problem, researcher conducted a simulation of selected spatial weighting matrix in conducting small area estimation by SEBLUP (Spatial Empirical Best Linear Unbiased Prediction) method and also pay attention to the many small areas. Spatial weighting matrix formation method that used can be classified into three major groups, weighting matrix based on geographical proximity, weighting matrix based on behavioral data and estimating the weighting matrix. The results of simulation studies generally show that the number of areas affecting the selection of optimum weighting matrix in the small area estimation. Small area of total (m = 16 area), spatial weighting matrix that is recommended is a spatial weighting matrix queen, spatial matrix weighting spatial K-NN, combination spatial weighting matrix K-NN and queen, spatial matrix weighting combination exponential and queen, and spatial weighting matrix combination radial and queen. Middle area of total (m = 64 area), spatial weighting matrix that is recommended the spatial weighting matrix spatial K-NN, spatial weighting matrix Radial, spatial weighting matrix combination Exponential and Queen, spatial weighting matrix combination K-NN and Queen, spatial weighting matrix combination power and Queen, spatial

weighting matrix combination double power and Queen and spatial weighting matrix combination Radial and Queen. While for the large area of the total (m = 144 area), spatial weighting matrix that is recommended is spatial weighting matrix Radial, spatial weightig matrix Queen, spatial weighting matrix combination Exponential and Queen, spatial weighting matrix combination power and Queen and spatial weighting matrix Radial and Queen.

Keywords : Small Area Estimation, SEBLUP, Spatial Weighting Matrix

1. INTRODUCTION

Model Fay and Herriot (1979), which became the basis for small area estimation assumes that the influence of area random error-free each others. But in some cases, this assumption is often violated. The reason is the diversity in an area influenced by the surrounding area, so that the spatial effect can be put into a random effect. The influence of spatial were common place between one area to another area, this means that one area affects other areas.

Model with consider the correlation spatial random effects within a small area estimation problem was first introduced by Cressie (1991) known as the SEBLUP (spatial Unbiased empirical best linear prediction) predictor. SEBLUP predictor was also used by Saei and Chambers (2003), Salvati (2004), Singh et al. (2005) and Pratesi and Salvati (2008). They inpute a spatial weighting matrix nearest neighbors into the model EBLUP.

Although previous researchers have explained about the spatial approach in a small area estimation, but there is a problem, that problem in determining the spatial weighting matrix that will be used in small area estimation. As explained by Getis and Aldstads (2004) that the spatial model, spatial weighting matrix is an important component in most models when the representation of the spatial structure is needed. Therefore it is necessary to do a special assessment regarding the optimum weighting matrix formation in a small area estimation. Stakhovych and Bijmolt (2008) and Jajang (2014) mentioned that the formation of spatial weighting matrix itself is classified into three major groups, based on geographical proximity, based on behavioral data and based on the prediction.

The method the researchers use in determining the optimum spatial weighting matrix is a literature study method and simulation methods. Simulations performed on several conditions to determine the optimum weighting matrix. Optimum weighting matrix obtained in the next simulation will be used in estimation of real data.

2. LITERATURE REVIEWS

2.1 Spatial Empirical Best Linear Unbiassed Prediction (SEBLUP)

Let defined vector $\boldsymbol{y} = (y_1, \dots, y_m)^T, \boldsymbol{v} = (v_1, \dots, v_m)^T$ dan $\boldsymbol{e} = (e_1, \dots, e_m)^T$, and matrix $\boldsymbol{X} = (\mathrm{x}_1^T, \dots, \mathrm{x}_m^T)^T$ and $\boldsymbol{Z} = diag(z_1, \dots, z_m)$. Based on the definition of vector and matrix, then the equation in matrix notation is:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \tag{1}$$

Model with spatial effect that is used in SAE model is *Simultaneously Autoregressive models* (SAR) that introduced by Salvati (2004), Pratesi dan Salvati (2008) dan Singh et al. (2005). Model SAR that introduced by Anselin (Chandra et al. 2007) where vector of area random effect $\boldsymbol{v}$ satisfy:

$$\boldsymbol{v} = \rho \boldsymbol{W} \boldsymbol{v} + \boldsymbol{u} \tag{2}$$

Equation (2) subtituted into equation (1) results:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{u} + \boldsymbol{e} \tag{3}$$

Prediction toward SEBLUP, by predictor EBLUP formulation SAR model is:

$$\hat{\theta}_i^{SEBLUP}(\hat{\sigma}_u^2, \hat{\rho}) = \boldsymbol{x}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{b}_i^T\{\hat{\sigma}_u^2[(\boldsymbol{I} - \rho \boldsymbol{W})(\boldsymbol{I} - \rho \boldsymbol{W}^T)]^{-1}\}\boldsymbol{Z}^T \times \{diag(\sigma_e^2) + \boldsymbol{Z}\hat{\sigma}_u^2[(\boldsymbol{I} - \rho \boldsymbol{W})(\boldsymbol{I} - \rho \boldsymbol{W}^T)]^{-1}\boldsymbol{Z}^T\}^{-1} \times \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \tag{4}$$

3. METHODOLGY

3.1 Simulation Studies

Simulation conducted to evaluates the good of the developed model. Simulation process conducted by following steps.

a. Make a artificial map square shaped of m small area arranged in the form of a square n x n area. Size m will be attempted is 16, 64 and 144. Each rectangular area with a size of 100 x 100.
b. Generating centroid coordinate data for each area based on a map created in step (a) and spread uniformly with maximum and minimum values follow the extents of each area.
c. Creating a spatial weighting matrix to be used based on the design map of the area and centroid created previously. The spatial weighting matrix to be formed are as follows:
    1. Distance Weighted Matrix
        - k-nearest neighbour matrix
        - radial distance matrix
        - power distance matrix

- eksponential distance matrix
- double power distance matrix

2. Weighted matrix based limit
   - Spatial contiguity weighted type queen
3. Limit and distance combination weighted matrix

d. Determine observation size in every small area

e. This simulation used one variable that became attention (y) and one concomitant variable x. Model used to achieved variable value that became attention (y) for small area to-i and unit to-j is as follow.

$$y_{ij} = \boldsymbol{x}_{ij}{}^{T}\boldsymbol{\beta} + v_i + e_{ij} \quad (6)$$

where x_{ij} is a concomitant variable, v_i is a area random effect, and e_{ij} is sample taking error .

1. Score $\boldsymbol{x}_{ij}$ generated normal spread with $\boldsymbol{x_{ij}} \sim N(3,5)$. Value $\boldsymbol{x_{ij}}$ that is obtained used to whole scenario in simulation process.
2. Determine $\boldsymbol{\beta} = (10\,,4)^T$ such that equation (20) became :

$$y_{ij} = 10 + 4\boldsymbol{x_{ij}} + v_i + e_{ij} \quad (7)$$

Score $\boldsymbol{v} = (v_1\,, \dots, v_m)^T$ generated by spread normal multavariat MVN(0,G) with $\boldsymbol{G} = \sigma_u^2\,[(\boldsymbol{I} - \rho\mathbf{W})(\boldsymbol{I} - \rho\mathbf{W}^T)]^{-1}$ is varians-covarians matrix *m x m* size with $\sigma_u = 3$, $\rho = 0.75$ and using $\mathbf{W}$ *queen*.

3. Score $\boldsymbol{e} = (e_{11}, \dots, e_{mN_m})^T$ generated spread normal with $\sigma_e = 1.34$

f. Calculate midle score of sample concomitant variable in each small area, by formulation:

$$x_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \quad (8)$$

g. Calculate midle score of variable considered for sample in each small area as direct predictor, by formulation:

$$y_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \quad (9)$$

h. Calculate varian score of variable considered by formulation:

$$s_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(y_{ij} - y_i\right)^2 \quad (10)$$

i. Find score $\hat{\theta}_i^{EBLUP}$ for model level area by using information y_i.

j. Find score $\hat{\theta}_i^{SEBLUP}$ untuk model level area by using information y_i.

k. Repeat step (d) until step (j) as many as B = 1000 such that can be calculated score of *relative root mean squares error* (RRMSE) from parameter predictor results in each area by formulation as follow:

$$RRMSE_{(i)} = \frac{1}{\theta_i}\sqrt{\frac{1}{B}\sum_{l=1}^{B}\left(\hat{\theta}_{il} - \theta_i\right)^2} \times 100\% \qquad (11)$$

l. Find score of *average relative root mean squares error* (ARRMSE)

$$ARRMSE = \frac{1}{m}\sum_{i=1}^{m} RRMSE_{(i)} \qquad (12)$$

m. Repeat step (d) until (l) for $\rho = 0.05, 0.25, 0.5$, $m = 64, 144$ and spatial weighting matrix **W** that is shaped on step (c) except step (c2), such that thenumber of scenario in this simulation is 132 scenario.

n. Repeat evaluation on all spatial weighting matrix by comparing score of *average relative root mean squares error* (ARRMSE) and predictor control $\rho = 0.05$ for each combination **W**, *m* and ρ.

3.2 Case Study

The case study in this study uses data of SUSENAS year 2010 and PODES 2011 issued by the Central Agency Statistics (BPS). The parameters observed in this study is that the average spending per household per month for sub-district in the city and district Bogor. Data that available on SUSENAS does not support direct estimation at the district level. This is because the sample at the district level is small. The model developed in this study is used as an alternative to overcome these problems. Modelling done by utilizing the information of the selected variables from the data of PODES as concomitant variables.

4. RESULTS AND DISCUSSION

4.1 Simulation for *m* = 16 area

Based on the results in Table 1, was obtained spatial weighting matrix details Recommended by taking into account the effect of correlation (ρ). In areas that have the effect of correlation area of strong (ρ = 0.75), then the matrix weighting spatial recommended is the matrix of weighted spatial K-NN, matrix weighting spatial rank of dual matrix weighting spatial combination of K-NN and queen and the matrix weighting spatial combination of double power and queen, to influence the correlation area under (ρ = 0.5), matrix weighting spatial recommended a matrix of weighting the

spatial queen, matrix weighting spatial combination of exponential and queen and the matrix weighting spatial combination of radial and queen, while for the effect of the correlation of low areas (ρ = 0.25), spatial recommended weighting matrix is queen spatial weighting matrix, the matrix exponential weighted combination of spatial and queen, and the spatial weighting matrix combination of radial and queen.

Table 1. Summary of Simulation Study Results score of ARRMSE for m = 16.

Spatial Weighted Matrix	ρ				
	0.05		0.25	0.5	0.75
	EBLUP	SEBLUP	SEBLUP	SEBLUP	SEBLUP
Exponential	9.46	6.72	6.65	6.56	6.42
K-NN	9.49	9.13	9.09	7.38	6.11
Power	9.46	9.21	9.09	7.35	6.62
Double Pwer	9.47	9.16	8.54	6.89	6.07
Radial	9.47	6.51	6.41	6.28	6.13
Queen	9.46	9.10	6.99	6.32	6.17
Exponential and *Queen*	9.48	9.07	7.00	6.28	6.15
K-NN and *Queen*	9.49	9.13	9.09	7.38	6.11
Power and *Queen*	9.46	9.23	9.08	8.86	6.55
Double Power and *Queen*	9.47	9.16	8.98	8.43	6.10
Radial and *Queen*	9.46	9.10	6.99	6.32	6.17

4.2 Simulation for m = 64 area

Table 2. Summary of Simulation Study Results score of ARRMSE for m = 64.

Spatial Weighting Matrix	0.05		0.25	0.5	0.75
	EBLUP	SEBLUP	SEBLUP	SEBLUP	SEBLUP
Exponential	7.05	6.02	6.03	5.95	5.74
K-NN	7.05	6.93	6.86	6.10	5.29
Power	7.05	6.60	6.04	5.91	5.65
Doble Power	7.05	7.06	6.93	6.34	5.63
Radial	7.05	6.64	6.43	6.09	5.80
Queen	7.05	6.80	6.64	6.18	5.54
Exponential and *Queen*	7.05	6.81	6.64	6.18	5.51
K-NN and *Queen*	7.05	6.91	6.70	6.11	5.40
Power and *Queen*	7.05	6.97	6.74	6.33	5.52
Double Power and *Queen*	7.05	7.07	7.01	6.28	5.61
Radial and *Queen*	7.05	6.80	6.62	5.96	5.47

Based on the results in Table 2, was obtained spatial weighting matrix details recommended by taking into account the effect of correlation (ρ). In areas that have the effect of correlation area of strong (ρ = 0.75), then the matrix weighting spatial recommended is the matrix of weighted spatial K-NN, to influence the correlation area under (ρ = 0.5), matrix weighting spatial recommended is the matrix of weighting the spatial combination of radial and queen, while for the effect of a low area correlation (ρ = 0.25), spatial recommended weighting matrix is a spatial weighting matrix rank.

4.3 Simulation for m = 144 area

Table 3. Summary of Simulation Study Results score of ARRMSE for m = 64.

Spatial Weighting Matrix	0.05		0.25	0.5	0.75
	EBLUP	SEBLUP	SEBLUP	SEBLUP	SEBLUP
Exponential	8.85	6.08	6.05	6.00	5.90
K-NN	8.84	8.87	8.83	7.62	6.15
Power	8.85	6.01	5.99	5.96	5.88
Double Power	8.85	8.62	8.58	7.51	6.14
Radial	8.85	8.60	8.03	6.71	6.14
Queen	8.84	8.62	8.31	6.75	6.03
Exponential and *Queen*	8.85	8.63	8.43	6.73	6.07
K-NN and *Queen*	8.84	8.86	8.82	7.61	6.15
Power and *Queen*	8.84	8.61	8.22	6.90	6.06
Double Power and *Queen*	8.85	8.63	8.58	7.66	6.13
Radial and *Queen*	8.85	8.63	8.43	6.91	6.06

Based on the results in Table 3, the spatial weighting matrix obtained details recommended by taking into account the effect of correlation (ρ). In areas that have the effect of correlation area of strong (ρ = 0.75), then the matrix weighting spatial recommended a matrix of weighting the spatial queen, matrix weighting spatial combination of exponential and queen, matrix weighting spatial combination of rank and queen, and the matrix weighting spatial combination of radial and queen, to influence the correlation area under (ρ = 0.5), matrix weighting spatial recommended a matrix of weighting the spatial radial matrix, weighted spatial queen and matrix weighting spatial combination of exponential and queen, while for the effect of the correlation of low areas (ρ = 0.25), spatial recommended weighting matrix is radial spatial weighting matrix.

4.4 Case Study

The data used in the study is SUSENAS year 2010 and PODES year 2011 issued by the Central Agency Statistics (BPS) for the city and district Bogor. The case of the same data has also been analyzed by Kurnia et al. (2015) by considering the influence of outliers with Winsor robust method SAE but did not consider spatial influences. Furthermore, Anisa et al. (2014) using the same data , also doing repair estimate in areas not surveyed in the SAE with a cluster analysis approach.

Based on data of SUSENAS and PODES, then the spatial weighting matrix is a matrix weighting for the number of areas being, namely the spatial weighting matrix power, spatial weighting matrix K-NN and spatial weighting matrix combination of radial and queen.

The variables observed in this study is the expenditure per household per month for sub-district in the city and distict Bogor. As for the auxiliary variables in matters of expenditure can be viewed from multiple proxies (approach) that is health and income. From the health side used variable number of families who receive the cards JAMKESMAS / JAMKESDA and the number of families living in neighborhood seedy. On the earnings used the number of number of family members of his family as farm laborers, the number of families who received a letter SKTM.

Furthermore, the variables considered (Y) will be seen the spatial dependence or in other words, whether or not there is a spatial autocorrelation. Measurement of spatial autocorrelation can be calculated using the Moran's Index (Moran), namely:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

To identify the existence of spatial autocorrelation or not, test of significance of Moran Index.

Table 4 Results of Spatial Autocorrelation test with Moran Indeks

Spatial Weighting Matrix	Indeks Moran	P.Value
K-NN	0.5151309	8.85E-10
Power	0.3732005	7.45E-05
Combination Radial and *queen*	0.3594768	4.47E-05

In Table 4 it can be seen that the spatial autocorrelation using the Moran index variables to consider (Y) for each spatial weighting matrix (W) produces a value P.Value smaller than at 0.05. Informsi Based on this, the spatial weighting matrix to be applied in the estimation SEBLUP on case study data are weighted spatial matrix K-NN, spatial weighting matrix power and spatial weighting matrix combination of radial and queen.

District and City of Bogor is composed of 46 districts and 428 villages/wards. There are two sub-district consisting of 14 villages were not studied because these districts didn't survey. So that approximately 25.93% of this amount or 111 villages / wards Susenas chosen as an example in 2010, the number of households for each village/urban neighborhoods chosen for example ranging from 15 to 107 households. The number of samples for each district is very small compared with the number of households in each of these districts, which ranged only between 12:14% and 12:22%.

After a thorough exploration of the data, small area estimation conducted by the direct method, EBLUP and SEBLUP, and the results can be seen in the following table.

Table 5 Small Area Estimation for the Average Expenditure Per Household SUSENAS 2010 Sub-District Level in the District and City of Bogor (Thousand)

Sub-District	Predictor				
	Direct	EBLUP	SEBLUP		
			K-NN	Power	Combination Radial and *Queen*
Nanggung	1212.44	1232.53	1254.7	1244.57	1234.57
Leuwiliang	1530.24	1518.48	1527.48	1496.78	1547.43
Pamijahan	1568.61	1555.73	1535.6	1549.04	1533.35
…	…	…	…	…	…
Parung panjang	515.82	530.55	550.67	546.73	548.17
Bogor selatan	5375.56	4894.54	4703.29	4784.95	5094.48
Bogor timur	10035.6	4573.3	5474.74	5378.02	5010.18

From Table 5 it can be seen that the entire predictor showed East Bogor districts have an average expenditure per household highest compared to other districts. Meanwhile, the district that allegedly has the lowest average expenditure per household is Parung panjang subdistrict.

Once the model is determined, then do a small area estimation and results to see a small area estimation can best be seen in Table 6

Table 6 Predictor ARMSE (*Average Root Mean Square Error*)

Predictor	ARMSE
Direct	603.428
EBLUP	442.58
K-NN	423.73
Power	430.488
Combination Radial and *queen*	433.596

To evaluate the the best estimation small area estimation method, can be seen in Table 6, which shows that the score predictor of Average Root Mean Square Error (ARMSE) on predictor SEBLUP with all matrix approach weighting spatial recommended smaller when compared with the estimate of direct and estimators EBLUP. So it can be said that the predictor SEBLUP better than the direct estimator or directly on SUSENAS 2010. Then in Table 6 it can be seen that the value estimator ARMSE on SEBLUP estimators using spatial weighting matrix K-NN is smaller than the predictor SEBLUP using matrix weighted spatial weighting matrix power and spatial combination of radial and queen, so it can be said that the predictor SEBLUP with spatial weighting matrix K-NN is better than the other estimators.

5. CONCLUSION AND SUGGESTION

5.1 Conclusion

Spatial weighting matrix is an important component in modeling spatial data where the data contained in spatial independence. Selection of different spatial weighting matrix will provide sensitive analysis results in many different types of studies. In addition, the determination of the weighting matrix is an important and fundamental thing in obtaining an accurate prediction. But the problem is determining the spatial weighting matrix is so large and yet their basic theory in determining the spatial weighting matrix the best at doing predictions at a certain condition.

The results of simulation studies generally show that many areas greatly affects the selection of spatial weighting matrix that can provide the best estimate. For small area

of total (m = 16 areas), spatial weighting matrix that is recomended is a spatial weighting matrix queen, spatial weighting matrix double power, spatial weighting matrix combination exponential and queen, and spatial weighting matrix combination of radial and queen. Furthermore, middle area of total (m = 64 areas), spatial weighting matrix that is recommended is a spatial weighting matrix K-NN, spatial weighting matrix power, and spatial weighting matrix combination of radial and queen. As for the large area of total (m = 144 area), spatial weighting matrix that is recommended is a radial spatial weighting matrix, weighted spatial queen, spatial weighting matrix combination exponential and queen, spatial weighting matrix combination power and queen and spatial weighting matrix radial and queen.

Then, for the case study for average household expenditure per capita level of sub-district in city or district Bogor in 2010 shows the results ARMSE estimator to estimate SEBLUP using spatial weighting matrix obtained in the simulation is smaller than the direct estimation and prediction EBLUP.

5.2 Suggestion

This research is still using the predictions for model of area level, so as for further research it is expected to be made to model the level of the unit and pay attention to the influence of spatial effects in its predictor.

REFERENCES

[1] Anisa R, Kurnia A and Indahwati. 2014. *Cluster information of non-sampled area in small area estimation*. IOSR Journal of Mathematics, 10(1), p: 15-19. doi: 10.9790/5728-10121519.

[2] Chandra H, Salvati N, Chambers R. 2007. *Small area estiamtion for spatially correlated populations a comparison of direct and indirect model-based methodes*. Statistics in transition 8:887-906.

[3] Cressie N. 1991. *Small area prediction of undercount using the general linear model, proceedings of statistics symposium 90: measurement and improvement of data quality*. Ottawa: Statistics Canada, pp. 93-105.

[4] Fay RE dan Herriot RA. 1979. *Estimates of income for small places an application of James-Stein procedures to census data.* Journal of the American Statistical Association 74: 269-277.

[5] Getis A dan Aldstadt J. 2004. *Constructing the spatial weights matrix using local statistic*. Geographical Analysis 36: 90-104.

[6] Jajang. 2014. *Modified local getis statistic on AMOEBA weights matrix for spatial panel model and its performance*. Bogor Agricultular University.

[7] Kurnia A. 2009. *An empirical best prediction method for logarithmic transformation model in small area estimation with particular application to susenas data*. Bogor Agricultular University.

[8] Kurnia A, Kusumaningrum D, Soleh AM, Handayani D, and Anisa R. 2015. *Small area estimation with winsorization method for poverty alleviation at a sub-district level*. International Journal of Applied Mathematics and Statistics™. 53 (6), pp. 77-84.

[9] Pratesi M and Salvati N. 2008. *Small area estimation: the EBLUP estimator based on spatially correlated random area effects*. Statistical methods and applications, Stat. Meth. & Appl. 17:113-141.

[10] Saei A and Chambers R. 2003. *Small area estimation: a review of methods based on the application of mixed models*. Southampton: Southampton Statistical Sciences Research Institute, WP M03/16

[11] Salvati N. 2004. *Small area estimation by spatial models: the spatial empirical best linear unbiased prediction (Spatial EBLUP).* Dipartimento di Statistica"G. Parenti" viale morgagni, 59-50134.

[12] Singh BB, Shukla K and Kundu D. 2005. *Spatial-temporal models in small area estimation*. Survey Methodology 31: 183-195.

[13] Stakhovych S, Bijmolt THA. 2008. *Specification of spatial models: a simulation study on weights matrices*. Papers in Regional Science. 88(2):389-408.