ON THE MODELLING OF LEPROSY PREVALENCE IN SOUTH SULAWESI USING SPATIAL AUTOREGRESSIVE MODEL[*]

Rezki Melany Sabil[1] and Ray Sastri2[2‡]

[1]Statistics Indonesia, Indonesia, rezkimelany@bps.go.id
[2]STIS-Polytechnics of Statistics, Indonesia, raysastri@stis.ac.id
[‡]corresponding author

Abstract

The prevalence of leprosy is the number of leprosy cases per 10000 peoples. Based on data from the Ministry of Health, the highest prevalence of leprosy was in South Sulawesi. This needs special attention because leprosy is a contagious disease. The number of leprosy cases in an area may be influenced by the number of leprosy cases in the neighbor area due to the movement of the air. So, the location of the area needs to be included in the analysis of leprosy. This study aims to identify the variables that spatially affect the prevalence of leprosy in South Sulawesi and modeling it. This study uses data from the Ministry of Health for the year 2016. The method of analysis is the Spatial Autoregressive Model (SAR). The result is there is a positive spatial autocorrelation in the prevalence of leprosy at the district level, which means that regions with a high prevalence of leprosy are surrounded by areas with a high prevalence of leprosy, and vice versa. The prevalence of leprosy in an area is influenced by the prevalence of leprosy in neighbor districts, the percentage of BCG vaccine recipient, and the percentage of households with a healthy lifestyle.

Keywords: leprosy, South Sulawesi, spatial autoregressive.

1. Introduction

The World Health Organization (WHO) in WHO (2015) stated that Indonesia was ranked first in Southeast Asia in the discovery of new cases of leprosy and the third highest in the world in 2016 after India and Brazil. This is certainly very alarming because this disease can spread to other people. Also, the sufferer experiences physical and psychological pain.

The prevalence of leprosy is the number of leprosy cases per 10.000 population. The higher the number, the bigger the problem that the government faces. Based on data from the Ministry of Health during 2008-2015, there were many new cases occurred. In 2010, the number of leprosy cases was 17.012, and rise to 20.160 cases in 2015. The case spread in all provinces in Indonesia with the highest prevalence was in Papua but the highest number of cases was in South Sulawesi. The prevalence of leprosy in South Sulawesi is always higher than the national prevalence.

The spread of leprosy is influenced by many factors. Some of them are the percentage of public health centers in an area (Ernawati et al., 2016), good housing sanitation (Ratnawati, 2016), and a clean lifestyle (Dzikrina & Purnami, 2013). Susanti & Azam (2016) did a deep survey in Pekalongan districts and found a relationship between BCG vaccination status, contact history, and duration of contact with the sufferer.

The spread of leprosy occurs through the air. The number of leprosy cases in an area may be influenced by the number of leprosy cases in the neighbor area due to the movement of the air. So, the location of the area needs to be included in the analysis of leprosy. Ernawati et al. (2016) have examined leprosy in East Java and found a spatial grouping of regions. Other research conducted by Noviani et al. (2014) got a conclusion that the Geographically Weighted Poisson Regression (GWPR) is the best model to explain the relationship between this disease and others the factors in Central Java. Fadmi (2015) also used GWPR when predicted of the number of new cases of leprosy in Buton District. There is not much research on leprosy in South Sulawesi, even though this province has a long history of leprosy.

The aim of this study is; (1) to identify the variables that spatially affect the prevalence of leprosy in South Sulawesi, (2) make a spatial model of the prevalence of leprosy in South Sulawesi.

2. Research Methodology

2.1 Data

This study uses data from the Ministry of Health for the year 2016. The unit of analysis is 24 districts with the list of variables presented in Table 1.

Table 1: The variables used in this study.

No.	Variable	Unit
1	The prevalence of leprosy (the ratio of sufferer per 10.000 population).	Ratio
2	Percentage of recipients of the BCG vaccine	Percentage
3	Percentage of Households with a healthy lifestyle	Percentage
4	The ratio of public health centers (puskesmas) per 100.000 population	Ratio
5	Percentage of decent sanitary houses	Percentage

2.2   Method of Analysis

Spatial analysis is an analytical method to obtain information about the effects of location. The description of spatial interactions between regions is manifested in the form of a weight matrix as follows:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NN} \end{bmatrix}$$

Rows and columns index the area. w_{ij} is the weight of the neighboring regions i and j, where $w_{ij}=1$ for the neighboring area and $w_{ij}=0$ for the other. One of the test statistics used to measure spatial autocorrelation is Moran's I with formula as follows:

$$I = \frac{n \sum_i^n \sum_{j \neq i}^n w_{ij} (y_i - \mu)(y_j - \mu)}{\sum_i^n \sum_{j \neq i}^n w_{ij} \sum_i^n (y_i - \mu)^2}$$

With μ is the average of variable Y. The *I* value will be in the interval -1 to +1. The positive value indicates that there is a grouping of an area with the same characteristics. And the negative value indicates that there is a grouping area with different characteristics. Values close to zero indicates that there is no spatial autocorrelation. The significance of Moran's I values is tested statistically with the Moran test.

The general model of spatial regression is as follows:

$$y = \rho W_1 y + X\beta + \lambda W_2 u + \varepsilon \tag{1}$$

with **y:** Vector of the response variable (nx1), **X**: Matriks of the explanatory variable (nxp), **β**: Vector of the coefficient of regression (px1), ρ: spatial lag coefficient, λ: spatial error coefficient, **u**,**ε**: vector of error (nx1), **W₁**: Weight matrix for spatial lag (nxn), **W₂**: Weight matrix for spatial error (nxn), n: number of observations.

The spatial general model above can be derived into three models:
1. The ordinary regression model (ordinary least square (OLS)), if ρ=0 dan λ=0,
   y = Xβ + ε.
2. Spatial autoregressive (SAR), if ρ≠0 dan λ=0, **y = ρW₁y + Xβ + ε.**

3. Spatial error model (SEM), if ρ=0 dan λ≠0 **y = Xβ + λW₂u + ε.**

The stage of analysis:
1. Make an OLS model.
2. Test the assumption for OLS (normality, multicollinearity, autocorrelation, heteroscedasticity).

3. Test the spatial heterogeneity using Breusch Pagan Test.   If spatial heterogeneity exists, then we use the GWR model. If it does not exist, we use OLS, SAR, or SEM.
4. Test the spatial dependency using Moran's I test. If Moran's I is significant, do the LM test dan RLM test to decide which model is the best; OLS, SAR, or SEM.

The statistics test for Moran's I test is:

$$z(I) = \frac{I - E(I)}{\sqrt{\text{Var}(I)}} \sim N(0,1)$$

$$E(I) = \frac{-1}{n-1}; Var(I) = \frac{n^2 \sum_i \sum_{j \neq i} w_{ij}^2 + 3\left(\sum_i \sum_{j \neq i} w_{ij}^2\right)^2 - n\ \sum_i \left(\sum_{j \neq i} w_{ij}\right)^2}{(n^2-1)\left(\sum_i \sum_{j \neq i} w_{ij}\right)^2}$$

We will be reject the null hypotheses if $|z(I)| > z\alpha/2$ or p-value $< \alpha$.
The parameters will be estimated using the Maximum Likelihood Estimation (MLE) method (Anselin, 1988). Selecting model will be carried out using the Lagrange Multiplier (LM) test and the Robust Lagrange Multiplier (RLM) test. LM test is used for comparing SAR and SEM to OLS while the RLM test is used for comparing SAR and SEM. The significance of the model is tested by the Likelihood Ratio Test and Wald Test. The Likelihood Ratio test is used to determine whether the spatial effects in the model obtained are significant or not, while the Wald test is used to determine the significant effect of each explanatory. In this study we use $\alpha = 0.10$.

Test for LM-Lag:
$\quad$ $H_0 : \rho = 0$; $H_1 : \rho \neq 0$

$\quad$ Test of Statistics: $LM_\rho \dfrac{\left[\frac{e\prime We}{\sigma^2}\right]^2}{T}$

$\quad$ With $D = \dfrac{(WXb)\prime M(WXb)}{\sigma^2}$

$\quad$ $T = tr[(W\prime + W)W]$, **b** is a vector of parameter estimator

Test for LM-Error:
$\quad$ $H_0 : \lambda = 0$; $H_1 : \lambda \neq 0$

$\quad$ Test of Statistics: $LM_\lambda = \dfrac{\left[\frac{e\prime Wy}{\sigma^2}\right]^2}{D+T}$

Test for RLM-Lag:
$\quad$ $H_0 : \rho = 0$; $H_1 : \rho \neq 0$

$\quad$ Test of Statistics: $RLM_\rho = \dfrac{\left[\frac{e\prime Wy}{\sigma^2} - \frac{e\prime We}{\sigma^2}\right]^2}{D}$

Test for RLM-Error:
$\quad$ $H_0 : \lambda = 0$; $H_1 : \lambda \neq 0$

$\quad$ Test of Statistics: $RLM_\lambda = \dfrac{(D+T)\left[\frac{e\prime We}{\sigma^2} - \left(\frac{T}{D+T}\right)\frac{e\prime Wy}{\sigma^2}\right]^2}{DT}$

The significance of the model is tested by the Likelihood Ratio Test and Wald Test. The Likelihood Ratio test is used to determine whether the spatial effects in the model obtained are significant or not. By testing for spatial lag as follows:

H_0 : $\rho_{model} = 0$ (no *spatial lag*)

H_1 : $\rho_{model} \neq 0$ (*spatial lag* is exist)

Test of Statistics: LR = -2S

With S is the difference in log-likelihood between ordinary regression and SAR.

The Wald test is used to determine the significancy effect of each explanatory.

H_0 : $\beta_k = 0$ (k-explanatory variable is not significant)

H_1 : $\beta_k \neq 0$ (k-explanatory variable is significant)

Test of Statistics: $Wald = (\frac{\beta k}{se(\beta k)})^2$

We reject the null hypothesis if p-value < α. In this study we use α=0.10

5. Make a weight matric, a spatial model, test the significance of the coefficient, and Interpreted.

3. Results And Discussion

The highest prevalence of leprosy is Sinjai District (3.3) and the lowest is Tana Toraja (0.04). It should be noted that there is a special hospital for leprosy in Tana Toraja, which affects the health services provided by lepers. Generally, the north area of South Sulawesi has a lower prevalence than the south.
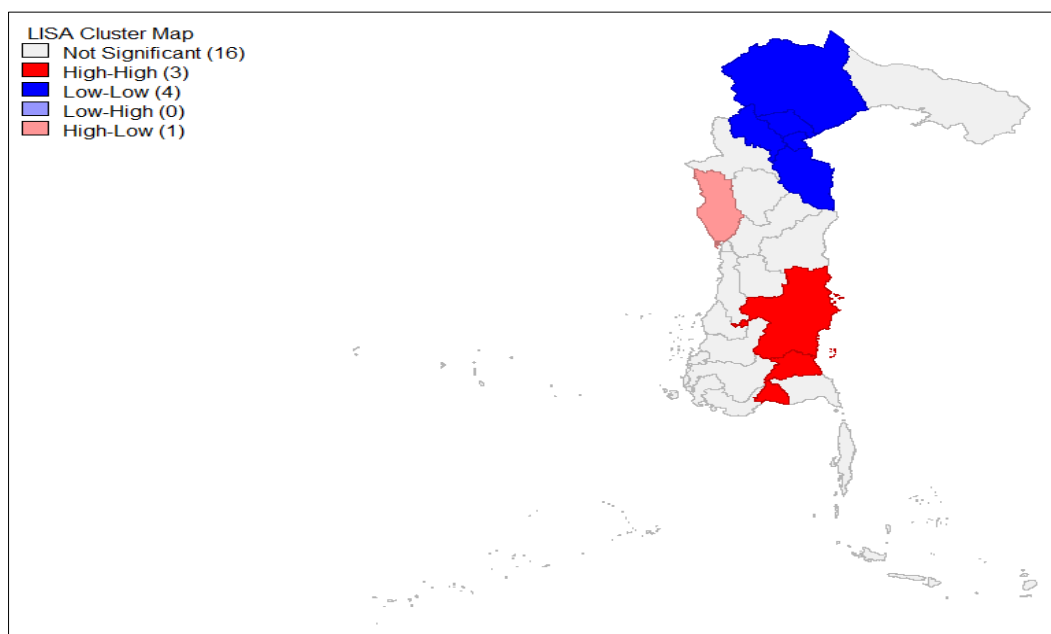

Figure 1: LISA *cluster map.*

Local Indicator of Spatial Association (LISA) cluster map in Figure 1 shows 8 districts with significant spatial autocorrelation, others are not significant. The red color

in the map describe areas with high prevalence surrounded by high prevalence. These areas include Bantaeng, Sinjai, and Bone. The blue one is describing low prevalence surrounded by low prevalence. These areas include Luwu, North Luwu, North Toraja, and Palopo. The pink one is for high prevalence surrounded by low. The region is Pinrang Regency.

Using a 10 percent significance level, the ordinary regression model shows that only the percentage of households with a healthy lifestyle significantly affects the prevalence of leprosy in districts level in South Sulawesi. The summary of the parameter can be seen in Table 2. Assumption of no autocorrelation is violated whereas the other assumption is met. To fix it, we tested the spatial heterogeneity.

Table 2: Summary of the statistical test in selecting spatial models.

Variable	OLS (R^2=0.254. AIC=64.59)		SAR (R^2=0.427. AIC=61.80)	
	Estimate	P-value	Estimate	P-value
(1)			(2)	(3)
Spatial Lag			0.461	0,015 *
Intercept	9.865	0.022*	7.643	0,014 *
BCG	-0.047	0.109	-0.038	0,081*
lifestyle	-0.029	0.060*	-0.023	0,040*
public health centers	-0.181	0.114	-0.137	0,112
sanitary	-0.015	0.267	-0.013	0,195

Based on Moran's Index test, there is significant and positive autocorrelation in the data. The moran index value is 0.4065. It means that there is a similarity in the prevalence of leprosy in the neighboring district. From the summary of the statistical test in selecting spatial models in Table 3, we conclude that the spatial autoregressive (SAR) model will be the best model for the leprosy data.

Table 3: Summary of the statistical test in selecting spatial models.

Test Statistics	df	value	p-value
LM-lag	1	5.161	0.023*
RLM-lag	1	5.036	0.025*
LM-error	1	2.412	0.120
RLM-error	1	2.286	0.131

Based on the likelihood ratio test, the SAR model is better than the OLS model. The R^2 in the SAR model is greater and the AIC is lower. Although the R^2 is greater but it too small (0.427). Another explanatory variable may include the model to improve performance. But unfortunately, the data about leprosy is very limited.

The prevalence of leprosy in neighbor districts, the percentage of BCG vaccine recipient, and the percentage of households with a healthy lifestyle have significantly affected the prevalence of leprosy in an area. The prevalence of leprosy in the neighbor area has significantly affected the prevalence of leprosy in a district. So, the

government policy to reduce the number of leprosy cases must consider the spatial effect. The percentage of BCG vaccines recipient and the percentage of households with a healthy lifestyle has a negative significant effect according to the parameter estimation. It means that the higher the percentage of BCG vaccines recipient, the lower the prevalence of leprosy in a district. The higher the percentage of households with a healthy lifestyle, the lower the prevalence of leprosy in a district. To reduce the leprosy case, BCG vaccines and a healthy lifestyle are very important. If the percentage coefficient of BCG vaccine recipient increases one percent, the prevalence will be reduced by 0.0384 people, or 4 people in one million people, assuming other variables are constant. If the number BCG vaccine recipient increases one percent, the prevalence of leprosy will be reduced by 0.0384 people, or 4 people in one million people, assuming other variables are constant. If the number of households with a healthy lifestyle increase one percent, the prevalence of leprosy will be reduced by 0.0233 people or 3 people in one million people.

The ratio of public health centers and the percentage of households with decent sanitary do not significantly affect the prevalence of leprosy in South Sulawesi. This variable may have not a direct effect on the prevalence of leprosy so that not significant. This variable is only a part of health services, the other variable can be included in the model in the next study. The housing condition in sanitation is not significant may be caused by the higher effect of a healthy lifestyle.

4. Conclusion

There is a positive spatial autocorrelation in the prevalence of leprosy at the district level in South Sulawesi in 2016, which means that regions with a high prevalence of leprosy are surrounded by areas with a high prevalence of leprosy, and vice versa. The prevalence of leprosy in an area is influenced by the prevalence of leprosy in neighbor districts, the percentage of BCG vaccine recipient, and the percentage of households with a healthy lifestyle. The best spatial model to describe the prevalence of leprosy in South Sulawesi is the SAR model.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Boston (US): Kluwer Academic.

Dzikrina, A. M., & Purnami, S. W. (2013). Pemodelan Angka Prevalensi Kusta dan Faktor yang Mempengaruhi di Jawa Timur dengan Pendekatan Geographically Weighted Regression (GWR). *Jurnal Sains Dan Seni*, *2*(2): D275–D281.

Ernawati, Latra, I. N., & Purhadi. (2016). Analisis Faktor-Faktor yang Memengaruhi Angka Prevalensi Penyakit Kusta di Jawa Timur dengan Pendekatan Spatial Durbin Model. *Jurnal Sains Dan Seni*, *5*(2): D295–D300.

Fadmi, F. R. (2015). Prediksi Jumlah Kasus Baru Kusta dengan Metode Geographically Weighted Poisson Regression (GWPR). *Jurnal Biometrika Dan Kependudukan*, *4*(1): 14–24.

Noviani, D., Wasono, R., & Nur, I. M. (2014). Geographically Weighted Poisson Regression (GWPR) Untuk Pemodelan Jumlah Penderita Kusta Di Jawa Tengah. *Jurnal Statistika*, *2*(2): 1–8.

Ratnawati, R. (2016). Faktor-Faktor yang Berhubungan Dengan Risiko Kejadian Penyakit Kusta (Morbus Hansen). *Tunas-Tunas Riset Kesehatan*, *6*(3): 103–109.

Susanti, K. N., & Azam, M. (2016). Hubungan Status Vaksinasi BCG, Riwayat Kontak, dan Personal Hygiene dengan Kusta di Kota Pekalongan. *Unnes Journal of Public Health*, *5*(2): 130–139.

[WHO] World Health Organization. (2015). Global leprosy update, 2014: need for early case detection. *Weekly Epidemiological Record= Relevé Épidémiologique Hebdomadaire*, *90*(36): 461–474.

APPENDIX 1. Spatial Weight

District	Number of Neighbors	Neighbor District	Weight of the neighbor district
Selayar	1	Bulukumba	0.4613
Bulukumba	3	Selayar, Bantaeng, Sinjai	0.1538
Bantaeng	4	Bulukumba, Jeneponto, Gowa, Sinjai	0.1153
Jeneponto	3	Takalar, Bantaeng, Gowa	0.1538
Takalar	3	Jeneponto, Gowa, Makassar	0.1538
Gowa	7	Bantaeng, Jeneponto, Takalar, Sinjai, Maros, Bonde, Makassar	0.0659
Sinjai	4	Bulukumba, Bantaeng, Gowa, Bone	0.1153
Maros	4	Gowa, Pangkep, Bone, Makassar	0.1153
Pangkep	3	Maros, Barru, Bone	0.1538
Barru	5	Pangkep, Bone, Soppeng, Sidrap, Parepare	0.0923
Bone	7	Gowa, Sinjai, Maros, Pangkep, Barru, Soppeng, Wajo	0.0659
Soppeng	4	Barru, Bone, Wajo, Sidrap	0.1153
Wajo	4	Bone, Soppeng, Sidrap, Lawu	0.1153
Sidrap	7	Barru, Soppeng, Wajo, Pinrang, Enrekang, Luwu, Parepare	0.0659
Pinrang	4	Sidrap, Enrekang, Tana Toraja, Parepare	0.1153
Enrekang	4	Sidrap, Pinrang, Luwu, Tana Toraja	0.1153
Luwu	7	Wajo, Sidrap, Enrekang, Tana Toraja, Luwu Utara, Toraja Utara, Palopo	0.0659
Tana Toraja	4	Pinrang, Enrekang, Luwu, Toraja Utara	0.1153
Luwu Utara	3	Luwu, Luwu Timur, Toraja Utara	0.1538
Luwu Timur	1	Luwu Utara	0.4613
Toraja Utara	4	Luwu, Tana Toraja, Luwu Utara, Palopo	0.1153
Makassar	3	Takalar, Gowa, Maros	0.1538
Parepare	3	Barru, Sidrap, Pinrang	0.1538
Palopo	2	Luwu, Toraja Utara	0.2306