www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-7, Issue-4, July-August 2017

International Journal of Engineering and Management Research

Page Number: 297-301

Conditional Autoregressive (CAR) Modeling Uses Weighted Matrix to First and Second Order (Case Study: Malaria Disease in Papua Province)

Siswanto[1], Muhammad Nur Aidi[2], Anik Djuraidah[3]
[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Science, Bogor Agricultural University, INDONESIA

ABSTRACT

Malaria is a contagious disease caused by plasmodium, a single-celled organism that belongs to the protozoa group. Malaria is transmitted through the bite of anopheles female mosquito that contains plasmodium in it. The occurrence or transmission of infectious diseases is determined by factors called host, agent, and environment. In Indonesia malaria is found to be widespread on all islands and Papua is the province with the highest API (Annual Parasite Incidence). This study employ publication data from the Central Bureau of Statistics and basic health research of Papua province with Conditional Autoregressive (CAR) method, which is one of the most commonly used models in analyzing areas especially in the field of epidemiology. Moran Index at first order shows that there is no spatial distribution pattern, while the second order there is spatial distribution pattern. The best Conditional Autoregressive (CAR) model is second order in determining factors affecting malaria disease in Papua because it has the lowest AIC value and variance. For second order significant factors are all explanatory variables namely the average of air temperature, air humidity, wind speed, rainfall, the proportion of households using such a mosquito net, mosquito coils, mosquito wire-netting, mosquito-prevention behavior using repellent, and malaria treatment with drug program by obtaining ACT drugs.

***Keywords*—** Malaria, Conditional Autoregressive (CAR), Second-Order, Moran Index

I. INTRODUCTION

Malaria is a contagious disease caused by plasmodium, a single-celled organism that belongs to the protozoa group. Malaria is transmitted through the bite of Anopheles female mosquito that contains plasmodium in it [1]. In Indonesia malaria is found to be widespread on all islands with varying degrees and weight of infections. Based on the API (Annual Parasite Incidence), Papua is the highest number of malaria incidents that is 42.64 percent. The occurrence or transmission of infectious diseases is determined by factors called host, agent, and environment [2].

Spatial aspect is evaluated importantly to be studied since inter-region certainly has different characteristics and problems. Spatial similarity is incorporated into the autoregressive model by putting on observations at other locations as addition covariates in the model with associated parameters defining spatial association. The autoregressive spatial models include Simultaneous Autoregressive (SAR) and Conditional Autoregressive (CAR). The SAR model is a spatial model observing random variables at one location simultaneously while the CAR model is a model monitoring random variables at any particular conditional location in neighboring locations [3], and CAR is often applied in analyzing areas especially in the field of epidemiology [4]. Through this study, it will be analyzed the significant factors affecting malaria disease in Papua Province by using Conditional Autoregressive (CAR) model with spatial weighted matrix to second-order.

II. METHODOLOGY

The data employed in this study are secondary data which are published by the Ministry of Health of the Republic of Indonesia with the title of Basic Health Research in Figure of Papua Province (2013) and The Central Bureau of Statistics of Papua Province (2013). The response and explanatory variables used can be seen in Table 1.

Table 1: Response and Explanatory Variables

Type of Variables	Name of Variables	Variables
Response	Malaria Sufferer (%)	Y
Explanatory	The average of air temperature (^{0}C)	X_1
	The average of air	X_2

humidity (%)	
The average of wind speed (knot)	X_3
The average of rainfall (mm)	X_4
The proportion of households using mosquito net (%)	X_5
The proportion of households using mosquito coils (%)	X_6
The proportion of households using mosquito wire-netting (%)	X_7
The proportion of households in mosquito-prevention behavior using repellent (%)	X_8
Malaria treatment With drug program by obtaining ACT drugs (%)	X_9

The analysis steps are as follows:

a. Data is clarified by using map to look over the spread of malaria disease in Papua Province.
b. Inspecting the distribution of data.
c. Forming first and second order spatial weighted matrix W by a queen contiguity method which is the value of 0 or 1 illustrating the structure of first and second-order neighborhood for each unit. If there are two adjacent areas, but between them there is one other area, then both are considered to have a second-order spatial neighborhood.
   i. Suppose that it will be created a spatial weighted matrix in Figure 1.

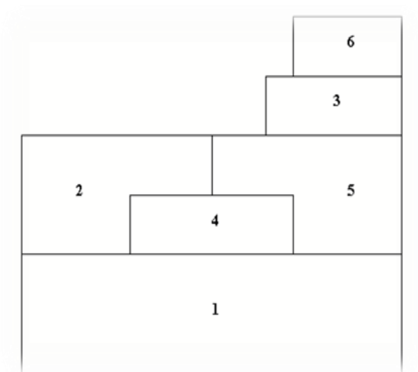

Figure 1: Illustration of Spatial Neighborhood

   ii. To first order, the matrix element would be 1 if the area to-*i* is adjacent to the region to-*j* and it would be 0 if the region to-*i* is not adjacent to the region to-*j*. According to the Figure 1, it is obtained W^* by the method of queen contiguity as follows:

$$W^* = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

   iii. In the second order, the matrix element would be 1 if the region to-*i* is adjacent even though between them there is one other region, both are said to have a second order spatial neighborhood with region to-*j*. According to the Figure 1, it is obtained W^* by the method of queen contiguity as follows:

$$W^* = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

d. If the spatial weighted matrix have been formed, then do the exploration of data to determine whether the spatial influence between districts/cities in Papua Province are exist or not by using Moran Index.
e. Testing the spatial correlation $(\hat{\rho})$.
f. Establishing the Conditional Autoregressive

III. OUR APPROACH

Data Description

In 2013 the number of API (Annual Parasite Incidence) in Papua reached 42.64%. Figure 2 illustrates that the district with the highest percentage of malaria sufferers is Intan Jaya which is labelled by a yellow color on the map, followed by Jayapura, Puncak Jaya, Mimika and Boven Digeol. On the other hand, the district with the lowest number of malaria sufferers is Nduga which is marked with blue on the map.

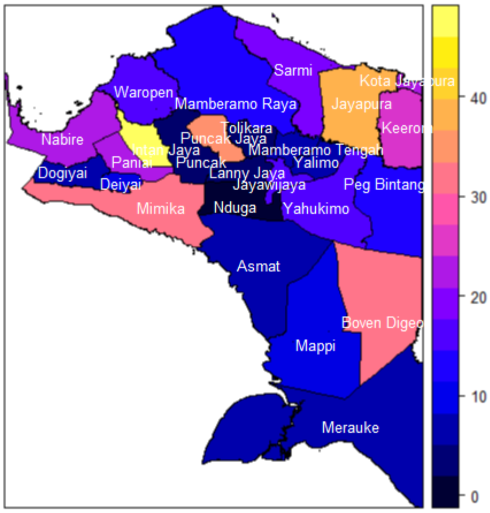

Figure 2: The Distribution of Malaria Sufferers in Papua Province

Spatial Data Exploration

Spatial Autocorrelation can be classified into two, namely the Global Association and Local Association. Global Association is the analysis of spatial association patterns on a wide scale to see the distribution of data. While Local Association is the quantification of spatial autocorrelation in smaller areas and it produces high statistical significance (hotspot), low statistical significance (coldspot) and outlier [5].

Moran Index Test

The Moran Index is employed to measure the global spatial autocorrelation which is interconnected spatially among locations. A Moran Index Testing is done by determining the weighted matrix first. The weighting matrix used is the weighted matrix Queen Contiguity where the weights based on the proximity of certain observational locations with other observational locations. The comparison of Moran Index values, expected values of *E(I)* and variance values of *Var (I)* are presented in Table 2. Based on the table, the Moran Index score for the percentage of malaria sufferers in Papua Province equal to $I = -0.093$ with p-value $0.679 (> \alpha = 0.05)$. This indicates that there is no spatial autocorrelation or pattern formed which is the random pattern in the transmitting of malaria in Papua Province on a first order spatial weighted matrix. For the second-order spatial weighted matrix, the Moran Index score for the percentage of malaria sufferers in Papua Province is $I = 0.107$ with the p-value $0.042 (< \alpha = 0.05)$. This points out that there is spatial autocorrelation in the transmission of malaria in Papua Province. A Moran index value which is more than $E(I) = -0.04$ indicate that there is a positive autocorrelation or clustering pattern and have similar characteristics of the location adjacent to the second-order.

Table 2: Comparison of Moran Index Value, Moran Index Expectation Value and Moran Index Variance Value

Order Weighted Matrix	Moran Index Value	E(I)	Var (I)	p-Value
First Order	-0.093	-0.0040	0.013	0.679
Second Order	0.107	-0.040	0.007	0.042

LISA (Local Indicator of Spatial Autocorrelation) Test

LISA testing provides various results for each district/city. The results of I_i and p-values are presented in Table 3 and Table 4. In Table 3, there are 3 districts that are significant at the level of $\alpha = 0.05$ such as Puncak Jaya, Mimika and Puncak. This indicates that globally there is no spatial autocorrelation at first order, but locally it indicates that the district has spatial relationships with its immediate neighboring districts.

Table 3: I_i Value and p-Value of LISA for Malaria Sufferers in Papua Province with First Order

No	Districts/cities	I_i	p-Value
1	Merauke	-0.428	0.400
2	Jayawijaya	0.544	0.633
3	Jayapura	0.249	0.581
4	Nabire	0.903	0.707
5	Paniai	1.081	0.709
6	Puncak Jaya	-5.840	0.002
7	Mimika	-5.275	0.018
8	Boven Digeol	-3.206	0.085
9	Mappi	0.368	0.614
10	Asmat	-0.565	0.434
11	Yahukimo	0.017	0.546
12	Peg. Bintang	-1.026	0.329
13	Tolikara	0.685	0.667
14	Sarmi	-0.025	0.528
15	Keerom	1.453	0.824
16	Waropen	-0.126	0.507
17	Mamberamo Raya	0.262	0.583
18	Nduga	1.038	0.714
19	Lanny Jaya	3.322	0.932
20	Mamberamo Tengah	2.839	0.901
21	Yalimo	-0.833	0.386
22	Puncak	-5.090	0.031
23	Dogiyai	-1.558	0.237
24	Intan Jaya	-0.476	0.436
25	Deiyai	-0.498	0.412
26	Jayapura City	1.022	0.788

In Table 4, there is a district that is real at the level of $\alpha = 0.05$ namely Puncak. So, Puncak, either in first order or second order indicates that there is spatial autocorrelation or spatial relationship with its immediate neighboring districts.

Table 4: I_i Value and p-Value of LISA for Malaria Sufferers in Papua Province with Second Order

No	Districts/cities	I_i	p-Value
1	Merauke	1.015	0.251
2	Jayawijaya	-0.331	0.491
3	Jayapura	-3.699	0.926
4	Nabire	-0.160	0.500
5	Paniai	-1.114	0.642
6	Puncak Jaya	3.699	0.073
7	Mimika	4.581	0.053
8	Boven Digeol	2.009	0.174
9	Mappi	0.099	0.447
10	Asmat	2.926	0.140
11	Yahukimo	0.034	0.449
12	Peg. Bintang	1.122	0.301
13	Tolikara	3.656	0.052
14	Sarmi	-0.214	0.476
15	Keerom	0.331	0.401
16	Waropen	0.018	0.451
17	Mamberamo Raya	-0.827	0.573
18	Nduga	1.214	0.308
19	Lanny Jaya	-2.910	0.809
20	Mamberamo Tengah	-0.816	0.595
21	Yalimo	2.224	0.189
22	Puncak	5.412	0.024
23	Dogiyai	0.690	0.342
24	Intan Jaya	0.913	0.315
25	Deiyai	0.153	0.436
26	Jayapura City	-0.411	0.551

Conditional Autoregressive (CAR) Model Analysis

Analysis of CAR model with first order weighted matrix in Papua Province involving administrative area without archipelago region using LRT test obtains spatial correlation value $\rho = 0.034$ with $LRT = 0.050$ and p-value =0.823. This shows an unreal model at a level of $\alpha = 0.05$ indicating that there is no spatial effect.

Analysis of CAR model with second order weighted matrix in Papua Province involving administrative area without archipelago region using LRT test obtains spatial correlation value $\rho = 0.491$ with $LRT = 508.44$ and p-value = 2.22E-16. This shows a real model at a level of $\alpha = 0.05$ indicating that there is no spatial effect.

Based on the results obtained, it can be seen that the spread of malaria disease in one region affects other regions for the weighted matrix of second order. The significance test of the variables in Table 5 reveals that all the variables included in the model are significant, i.e X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, and X_9.

According to table 5, the increase in the average of air temperature (X_1), wind speed (X_3), Rainfall (X_4), X_5, X_6, and X_9 around the object or location about one unit will lead the percentage of malaria patients to go up by 0.295, 0.532, 0.340, 0.132, 0.401, and 0.347 percent respectively. However the rise in the average of air humidity (X_2), X_7, and X_8 around the object or location by one unit will stimulate the percentage of malaria sufferers to decline approximately 0.721, 0.115, and 0.286 percent successively.

Table 5: The Comparative Analysis of CAR

	Order			
	First-Order		Second-Order	
	Coefficient	p-Value	Coefficient	p-Value
ρ	0.034	0.823	0.491	2.2E-16
Intercept	0.004	0.980	0.184	2.2E-16
X_1	-0.085	0.733	0.295	2.2E-16
X_2	0.418	0.059	-0.721	2.2E-16
X_3	-0.338	0.047	0.532	2.2E-16
X_4	0.347	0.068	0.340	2.2E-16
X_5	-0.011	0.963	0.132	2.2E-16
X_6	-0.092	0.702	0.401	2.2E-16
X_7	-0.442	0.193	-0.115	2.2E-16
X_8	0.334	0.296	-0.286	2.2E-16
X_9	0.872	3.02E-05	0.347	2.2E-16

Some of the criteria used in viewing the model goodness of fit test in the CAR are AIC (Akaike's Information Criterion), variance estimator and spatial correlation coefficient value [6]. Table 6 illustrates the goodness of fit test of CAR model with second order is better than CAR model weighted first-order. Malaria sufferer with second-order shows smaller value of $AIC = -465.1$ than first-order. Perceived from the value estimator variance of a CAR model of Second order is smaller than the first order of CAR model that is $5.4781E-10 < 1.7086E-09$.

Table 6: The Comparative Analysis of Model Goodness of Fit on CAR

Criteria	First-Order	Second-Order
AIC	71.687	-436.71
σ^2	0.59384	1.7086E-09
ρ	0.034572	0.491

IV. CONCLUSION

Some conclusions that can be concluded in this research area are:

a. Based on Moran Index there is a spatial distribution pattern of malaria disease in second order. Spatial correlation on Conditional Autoregressive (CAR) model has a significant p-value which is able to reveal the relationship among regencies/cities in Papua Province in second-order while spatial correlation by using first order is insignificant.

b. The best CAR (conditional autoregressive) model is the second order in determining the factors which affect malaria disease in Papua Province since it has the smallest of AIC value and variance. Significant factors at first order are the average wind speed and treatment of malaria with drug program by getting ACT drugs. For second order, the significant factors are all explanatory variables namely the average of air temperature, air humidity, wind speed, rainfall; the proportion of households using such a mosquito net, mosquito coils, mosquito wire-netting, mosquito-prevention behavior using repellent, and malaria treatment with drug programs as ACT medication.

REFERENCES

[1] Ministry of Health of Republic of Indonesia. “Malaria.” Indonesia. ISSN 2442-7659, Apr. 10, 2016.
[2] A.A. Arsin. *Malaria di Indonesia Tinjauan Aspek Epidemiologi*. Masagena Press, 2012, pp. 85-100.
[3] L.A. Waller and C.A. Gotway. *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons, 2004, pp. 362-379.
[4] V.D. Oliveira, “Bayesian analysis of conditional autoregressive models.” *Ann Inst Stat Math*, vol. 64, pp. 107-133, 27 May 2010.
[5] J. Lee and S.D. Wong. *Statistical Analysis with Arcview GIS*. John Wiley & Sons, 2001, pp. 132-177.
[6] M. Meilisa. “Model Otoregresif Simultan dan Otoregresif Bersyarat untuk Analisis Kemiskinan di Provinsi Jawa Timur.” M.Si. thesis, Bogor Agricultural University, Bogor, 2011.