

# Rice Production Modelling in Indonesia with the Spatial Autoregressive (SAR) Approach

Tuti Purwaningsih<sup>1,\*</sup>, Nur Hargita Padmarini<sup>2</sup>

<sup>1</sup>*Department of Statistics, Universitas Islam Indonesia*

<sup>\*</sup>*Corresponding author: tuti.purwaningsih@uii.ac.id*

## ABSTRACT

Rice production basically depends on the variable harvested area and yields per hectare, rice production can be increased if the harvested area has increased. In this study, a description of rice production and the factors that influence it from a regional perspective with a spatial weighting matrix, and modelling of rice production using the Spatial Autoregressive. The results showed that the spread of rice production in Indonesia has a pattern of clustering between regions that are close to one another. Based on the relationship of rice production with the variables that influence it, namely rice harvest area, paddy productivity, number of rice farmers, and paddy land area, it can be interpreted that similarities and differences in characteristics in each of the adjacent provinces can lead to an increase or decrease in rice production in Indonesia. The model used is the SAR (Spatial Autoregressive) model, and the results show that rice production in an area is influenced by rice harvested area and paddy field area.

**Keywords:** *Rice Production; Crops; SAR; Clustering; Indonesia.*

## 1. INTRODUCTION

Food is one of the main human needs, and its sufficiency is part of human rights guaranteed in the law as a basic component to reach quality human resources. Food availability, according to [5] is determined by 3 main aspects, namely production (quantity), distribution (accessibility), and consumption (nutritious and safe).

Rice is one of the most important staple food products in the world. This statement is especially true in Asia, where rice is the staple food for the majority of the population (especially among the middle to lower classes of society). Asia is also home to farmers who produce around 90% of total world rice production, amounting to 115.5 kg / capita / year. Indonesia is one of the countries in Asia that produces the most rice. The continued dominance of rice consumption, of course, presents even greater challenges for efforts to increase food security through increased rice production. Especially with the emergence of the threat of decreased production due to the increasing number of agricultural lands that have changed functions.

Tobler in 1970 stated the first law about geography, namely the condition at one point or area associated with conditions at one of the adjacent points or areas. This law is the basis for regional scientific studies. Spatial effects often occur between one region to another. In spatial data, often observations at one

location depend on observations at other nearby locations (neighboring).

## 2. LITERATURE REVIEW

### 2.1. Research Literature

According to the Ministry of Agriculture ([www.pertanian.go.id](http://www.pertanian.go.id)), from 2011 to 2017 the increasing trend of rice production also continued to increase, namely 65.75 million tons in 2011 and 81.38 million tons in 2017. The achievements in 2017 actually exceeded the rice production target set at 79 million ton, making 2.56% of the previous year's growth.

The increase in production over the past 10 years, especially in recent years is the result of the ~agricultural machine tools, ponds, and price guarantees for farmers.

In terms of rice consumption, the trend always follows population growth every year. Based on the Central Statistics Agency (BPS) 2018, the rice surplus in 2017 is 13.81 million tons.

### 2.2. Spatial Regression Model

According to Tobler's first law [3] said that "Everything is related to everything else, but near things are more related than distant things". That law is the reference for regional studies. Spatial effects are usually found in one area to another, meaning that observations in an area depend on observations in neighboring areas or adjacent areas. Spatial

regression is a statistical method used to determine the relationship between response variables and predictor variables by considering inter-regional linkages. Anselin developed a general model of

$$y = \rho W_1 + X\beta + u, \text{ where } u = \lambda W_2 + \epsilon \quad (1)$$

$$y = (I_N - \rho W_1)^{-1} X\beta + (I_N - \rho W_1)^{-1} (I_N - \lambda W_2)^{-1} \epsilon$$

$$\epsilon \sim N(0, \sigma^2 I_N)$$

spatial regression using crossed data [1]. The general model of the spatial regression equation is written in the form of:

### 2.3. Moran Index

Globally testing through Moran's I statistics is an autocorrelation test assuming the same location but different variables and covariance based. According to Lee and Wong (2001) Moran's I statistics can be measured by the following formula.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

$\bar{x}$  and  $\bar{y}$  the Pearson correlation equation is an average sample of predictor variables and the response. P value is used to measure whether the predictor variables and the response correlated.

According to Lee and Wong (2011), "The coefficient of Moran's I used to test the spatial dependency or autocorrelation between observations or location". Hypothesis is:

H0: I = 0 (no autocorrelation between locations)

H1: I ≠ 0 (no autocorrelation between locations)

The test statistic used is as follows:

$$Z_{hitung} = \frac{I - I_o}{\sqrt{\text{var}(I)}} \sim N(0,1) \quad (3)$$

where :

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

$$E(I) = I_o = -\frac{1}{n-1}$$

$$\text{var}(I) = \frac{n^2 S_1 - n S_2 + 3 S_o^2}{(n^2 - 1) S_o^2}$$

$$S_1 = \frac{1}{2} \sum_{i \neq j}^n (w_{ji} + w_{ij})^2$$

$$S_2 = \sum_{i=1}^n (w_{io} + w_{oi})^2$$

$$S_o = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

$$w_{io} = \sum_{j=1}^n w_{ij}$$

$$w_{oi} = \sum_{j=1}^n w_{ji}$$

Information:

$x_i$  = variable data to location-i (i = 1, 2, ..., n)

$x_j$  = variable data to location-j (j = 1, 2, ..., n)

$\bar{x}$  = Average Data

var (I) = varians Moran's I

E(I) = expected value Moran's I

Decision-making reject Ho if  $|Z_{hitung}| > Z_{\alpha/2}$ .

The value of the index I is between -1 and 1. If I > I<sub>o</sub>, the data has a positive autocorrelation, if I < I<sub>o</sub>, the data has negative autocorrelation, and Moran index value is zero indicating no groups. Moran index value does not guarantee the accuracy of measurement if

the weighting matrix used are not standardized weighting.

### 3. METHODS

The data used are secondary data, taken from the website of the Ministry of Agriculture of the Republic of Indonesia in 2017. Variables used are: Rice production, Rice harvested area, rice productivity, number of rice farmers, and rice field area. The method of analysis used is the spatial regression analysis method, namely the Spatial Autoregressive Model (SAR). The research method is carried out in steps such as the following:

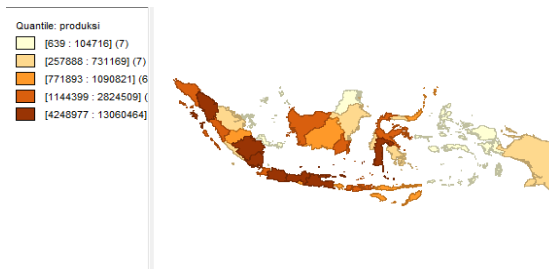
1. Exploring data on thematic maps to find out the distribution patterns and spatial dependencies on each variable.
2. Perform regression modelling with the Ordinary Least Square (OLS) method which includes parameter estimation, estimation of the significance of the model, residual test (identical, independent and normally distributed)

3. Determine the spatial weighting matrix.
4. Test spatial dependencies or spatial correlations between observations that are adjacent to the Moran or Morans' index.
5. If there is a spatial dependency, then the Spatial Autoregressive (SAR) modelling will then be analyzed.
6. Summing up the results obtained.

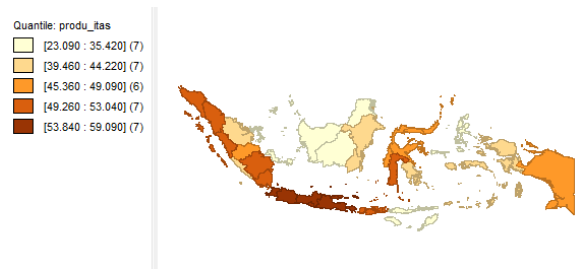
### 4. RESULT AND DISCUSSION

#### 4.1. Thematic Map

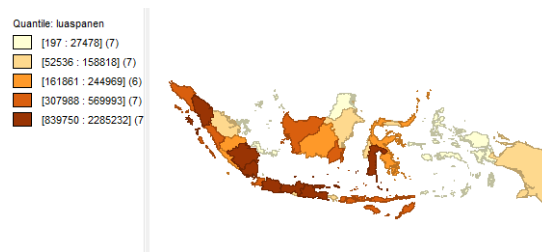
On the thematic map in Figure 2, there is a color degradation that shows the value of the sum of each variable. The more concentrated the color of the region, the higher the number of each variable in the region. To find out the number of each variable, it is categorized into five categories, namely very low, low, medium, high and very high.



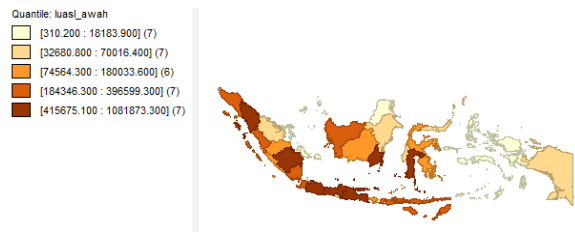
(a) Map of total rice production



(b) Rice productivity map



(c) Map of rice harvested area



(d) Extensive paddy field map



(e) Map of the number of rice farmers

Figure 1. Thematic Map

Based on the mapping results in Fig 1., it shows that there is a pattern of spatial correlation between regions. This indicates that there is a positive spatial correlation between regions, because the same colors are closed to each other; it explains that the same values are also closed to each other. This analysis is not necessarily true before the study continues by calculating the Moran Index between variables.

Rice field area	0.297096
Number of rice farmers	0.508156

4.1.1 Moran Index

Table 1. Moran Index

Variable	Moran Index
Rice Production	0.410254
Rice Productivity	0.607216
harvested area	0.36529

From the calculation of Moran index, it was found that all variables get a positive Moran index, namely rice production with a value of  $I = 0.41$ , rice productivity with a value of  $I = 0.607$ , the harvested area with a value of  $I = 0.365$ , paddy area with a value of  $I = 0.297$ , the number of rice farmers with a value of  $I = 0.508$ . A positive Moran index value indicates that the variable has an element of closeness or closeness between the interconnected regions.

4.1.2. SAR Model

After calculating the Moran index value, the next analysis uses the SAR Model. Following is the output of the SAR model:

Variable	Coefficient	Std.Error	z-value	Probability
W_produksi	0.00550235	0.0149806	0.367299	0.71340
CONSTANT	-484649	207696	-2.33345	0.01962
produ_itas	8044.33	4915.26	1.6366	0.10171
luaspanen	7.19635	0.437468	16.45	0.00000
luasl_awah	-4.00093	0.757759	-5.27995	0.00000
jumlah_tani	0.297905	0.148408	2.00734	0.04471

Figure 2. Output SAR all variables

From the output in Figure 3. there are insignificant variables, namely rice productivity and the number of farmers. After removing the insignificant variables,

get the output with all the significant variables as follows:

Variable	Coefficient	Std. Error	z-value	Probability
W_produksi	0.0212641	0.0144708	1.46944	0.14171
CONSTANT	-162940	56311.4	-2.89355	0.00381
luaspanen	7.82288	0.373973	20.9183	0.00000
luasl_awah	-4.66522	0.762336	-6.11965	0.00000

Figure 3. Output SAR significant variables

Based on the above output, the Spatial Autoregressive (SAR) Model for Rice Production in Indonesia in 2017 can be obtained as follows.

$$y_i = 0.0212641 \sum_{j \neq i}^N W_{ij} y_j + 7.82288 X_2 - 4.66522 X_3 + \varepsilon_i$$

At the significance level  $\alpha = 5\%$ , it can be seen that Rice Production in area  $i$  is influenced by Rice Production from their neighbors. The interpretation of the Spatial Autoregressive model is for each increase of 1% of the Rice Harvest Area will increase the average Rice Production by 100 (e7,82288-1)%. The Rice Field Variable has a negative regression coefficient so for every 1% increase in the Rice Field Area will decrease Rice Production by 100 (1 - e-4,66522 (1))%.

### 5. CONCLUSION

From the analysis, it can be concluded that the distribution of rice production in Indonesia has a clustered pattern between regions that are close to one another. Based on the relationship between Rice Production and the variables that influence it, namely Rice Harvest Area, Rice Productivity, Number of Rice Farmers, and Area of Rice Fields, it can be interpreted that similarities and differences in characteristics in each of the adjacent Provinces can lead to an increase or decrease in rice production in Indonesia. Rice production data in Indonesia in 2017 can be modelled using SAR and there are two factors that significantly affect Rice Production, namely Rice

Harvest Area and Rice Area. In addition, this research can also be used as a perspective for the government in Indonesia to increase National Income by increasing the average value of Rice Production.

### REFERENCES

- [1] Anselin, L. (2013). Spatial Econometrics: Methods and Models (Vol. 4). Springer Science & Business Media.
- [2] Lee, J., Wong, D. W. (2001). Statistical Analysis with ArcView GIS. United States: John Wiley & Sons.
- [3] Miller, H. J. (2004). Tobler's First Law and Spatial Analysis. Annals of the Association of American Geographers. 94(2): 284-289.
- [4] [Muhardi, R. (2018). Analisis Spasial dengan Open Source GIS. Yogyakarta: Mobius.
- [5] Pujiasmanto, B. (2014). Perkuat Ketahanan Pangan Nasional Kita. Fakultas Pertanian. UNS. Surakarta
- [6] Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography. 46: 234-240.
- [7] Ministry of Agriculture. 2018.