Spatial autocorrelation analysis of Covid-19 cases in South Kalimantan, Indonesia

O N Amaliah*, Y Sukmawaty, and D S Susanti

Statistic Study Program, Faculty of Mathematics and Natural Science, Lambung Mangkurat University, Indonesia

*oktaviaamaliana@gmail.com

Abstract. Coronavirus Disease 2019 (COVID-19) is a new coronavirus that was discovered in Wuhan, China, at the end of 2019. In March 2020, the outbreak extended throughout the world, including Indonesia and one of its provinces, South Kalimantan. This rapid expansion should be linked to people's mobility between regions, hence the linkage across regions must be examined. In South Kalimantan Province, the purpose of this research is to evaluate the distribution and relationship across regions in terms of the number of positive COVID-19 cases, the number of additional positive COVID-19 cases, and the number of COVID-19 patients under treatment. The spatial autocorrelation analysis with the Moran Index and Local Indicator of Spatial Autocorrelation (LISA) tests were used to determine the spatial autocorrelation between what and what using what data/where the data obtained? from March 22 to September 30, 2020. Based on the results of the Moran Index test, it is known that there is a spatial autocorrelation in the number of cases, the number of additional cases and the number of positive confirmed COVID-19 patients in treatment between one region and another neighboring area. While the results of the LISA Index test show that Balangan Regency, Hulu Sungai Tengah Regency, Hulu Sungai Utara Regency, Banjarmasin City, Tabalong Regency and Banjar Regency affect the level of COVID-19 cases in their respective neighboring areas. Therefore, there is a need for policies to control community mobility in those spatially correlated areas and increase testing and tracing to control the spread of COVID-19 cases in South Kalimantan Province.

1. Introduction

Coronavirus Disease-2019 (COVID-19) is a new outbreak caused by Coronavirus-severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) [1]. It first occurred in Wuhan City, Hubei Province, China in December 2019. Since the initial confirmation, in March 2020 [2], the virus's spread has accelerated, notably in Indonesia. In South Kalimantan Province, the number of positive cases of COVID-19 cases has increased from March to September [3]. There is a need to prevent further increase in the cases of COVID-19 in South Kalimantan Province. In China, prior research on the distribution of COVID-19 was conducted utilizing a spatial autocorrelation technique using the Moran Index test [4]. The findings showed that cases in one part of China affect other parts of the country, indicating that the COVID-19 pandemic is spreading spatially. Another research has utilized the LISA Index test to find the impact of an epicenter area in UK on its adjacent places due to community movement [5]. Given the importance of spatial (regional) understanding of COVID-19 spread, this research aimed to analyze whether there is a relationship across regions in terms of the
number of cases, the number of additional cases, and the number of patients who were in the treatment of Covid-19 in South Kalimantan Province using a spatial autocorrelation approach. The results of the analysis could be used as a baseline in making evidence-based policies and useful information for public health.

2. Deep learning theory

2.1. Spatial Autocorrelation
The estimation of the correlation between a variable and itself in a space (distance, time, and region) is known as spatial autocorrelation [6]. Figure 1 (a) indicates that there is a positive spatial autocorrelation when two areas that are adjacent to each other have comparable characteristics. However, if two areas that are adjacent to each other have different characteristics, then there is a negative spatial autocorrelation, as shown in Figure 1. (b). Figure 1 (c) on the other hand, has random properties, indicating that there is no spatial autocorrelation.

![Figure 1. Spatial Autocorrelation](image)

2.2. Moran Index
The Moran Index is a variation of the Pearson correlation that is used to determine the spatial relationship between two areas. The Moran Index determines the spatial correlation of a single variable, such as \(x_{ij}\) (\(x_{12}\) and \(x_{13}\)) [7]. The Moran Index is calculated using the following formula:

\[
I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left( x_{i} - \bar{x} \right) \left( x_{j} - \bar{x} \right)}{\sum_{i=1}^{n} \left( x_{i} - \bar{x} \right)^{2}}
\]

Where:
- \(I\) = Moran Index
- \(n\) = Number of observation (area)
- \(\bar{x}\) = Mean of variable
- \(x_{i}\) = Observation value at location-\(i\)
- \(x_{j}\) = Observation value at location -\(j\)
- \(w_{ij}\) = Elements of the weighted matrix between location-\(i\) and location-\(j\)

The spatial pattern can be determined using the Moran Index value criteria as follows: if \(I > E(I)\), then the data has a clustered pattern; if \(I = E(I)\), then the data has a random pattern, and if \(I < E(I)\), then the data has a spread pattern, where \(E(I)\) can be obtained using the following formula:

\[
E(I) = \frac{-1}{(n - 1)}
\]

Where:
- \(E(I)\) = Moran Index expectation
- \(n\) = Number of observation (area)
the hypotheses of the Moran Index are as follows:

\( H_0 : I = 0 \) (There is no spatial autocorrelation between locations)

\( H_1 : I \neq 0 \) (There is a spatial autocorrelation between locations)

the statistical test of Moran Index is calculated as follows:

\[
Z_{hitung} = \frac{I - E(I)}{\sqrt{\text{var}(I)}}
\]  

(3)

Where the Moran Index's variance value is:

\[
\text{Var}(I) = \frac{n^2S_1 - nS_2 + 3S_0^2}{(n^2 - 1)S_0^2} - [E(I)]^2
\]  

(4)

\[
S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}
\]  

(5)

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

(6)

\[
S_2 = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji} \right)^2
\]  

(7)

The null hypothesis, \( H_0 \), of Moran Index is rejected if \( Z_{count} > Z_{\alpha} \), indicating that there is a spatial autocorrelation between regions.

2.3. Moran’s Scatterplot

Moran's Scatterplot consist of \( Z_{count} \) of the Moran Index for the location and the \( Z_{count} \) on the neighbouring area of observation for the location. The area may be seen in the Moran's scatterplot quadrant, depending on the observed value and the average value [6]. A high observation area surrounded by other high observation areas is represented in Quadrant I (High-High). A low observation region but it is surrounded by a high observation area is shown in Quadrant II (Low-High). A low observation area surrounded by other low observation areas is represented in Quadrant III (Low-Low). A high observation area surrounded by the low observation area is shown in Quadrant IV (High-Low).

![Figure 2. Moran’s Scatterplot](image)

2.4. Local Indicator of Spatial Autocorrelation (LISA) Index

The Local Indicator of Spatial Autocorrelation (LISA) Index can be used to identify local spatial autocorrelation between two observation areas. The index can be calculated using the following formula:

\[
L_i = \frac{(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2/n}
\]  

(8)

The hypotheses to test the LISA Index are as follows:
\(H_0: I = 0\) (At area \(i\) there is no spatial autocorrelation.)

\(H_1: I \neq 0\) (At area \(i\) there is spatial autocorrelation.)

The statistical test of the Moran Index is calculated as follows:

\[ Z_{hitung} = \frac{L_i - E(L_i)}{\sqrt{\text{var}(L_i)}} \] (9)

Where the LISA Index's variance value is:

\[ \text{Var}[L_i] = w_i^{(2)} \left( \frac{n - m_s}{m_s^2} \right) + 2w_i^{(kh)} \left( \frac{2m_s}{m_s^2} - n \right) \left( \frac{2m_s}{m_s^2} - n \right) - [E(L_i)]^2 \] (10)

\[ w_i^{(2)} = \sum_{j=1}^{n} w_{ij}^2 \] (11)

\[ m_s = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \] (12)

\[ 2w_i^{(kh)} = \sum_{k \neq l} \sum_{h \neq i} w_{ik}w_{ih} \] (13)

The null hypothesis, \(H_0\), of the LISA Index is rejected if \(Z_{\text{count}} > Z_{\alpha}\), indicating that at position \(i\) there is spatial autocorrelation.

2.5. Spatial weights matrix

The standardized contiguity matrix is used to generate the spatial weighting matrix \((W)\) [6]. The contiguity matrix's elements are worth 1 if the observed area is directly adjacent to the neighboring region, and 0 if it is not. The contiguity matrix takes the following generic form:

\[ C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \] (14)

The procedure of standardizing contiguity matrix is to make the number of rows equal to one [6]. The values in the standardized matrix are weighted as follows:

\[ w_{ij} = \frac{c_{ij}}{\sum_{j=1}^{n} c_{ij}} \] (15)

A standardized spatial weighting matrix can be formed as follows:

\[ W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{11} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \] (16)

2.6. Thematic Map

Thematic maps represent the usage of space in the observed area, including both qualitative and quantitative data distributed geographically. Population density, climate, and health issues can all influence geographic distribution. Color gradations depending on the level of observational data are used to distinguish thematic maps [8].

3. Method

The secondary data was derived from the South Kalimantan Provincial Health Office's Instagram. The research data was gathered from the South Kalimantan Provincial Health Office's publication, which ran from March 22nd, 2020 to September 30th, 2020. This research has involved 13 districts/cities in the province of South Kalimantan. The research variables are described in Table 1.
### Table 1. Research Variables

| Variable                                                      | Symbol | Unit   | Data Scale |
|---------------------------------------------------------------|--------|--------|------------|
| Number of positive COVID-19 cases                             | $X_1$  | Person | Ratio      |
| Number of additional positive COVID-19 cases                  | $X_2$  | Person | Ratio      |
| Number of COVID-19 patients who were under treatment          | $X_3$  | Person | Ratio      |

The following sections describe the various stages of the research:

- a. Describe the statistical characteristics of each research variable
- b. Create a Queen Contiguity type spatial weights matrix;
- c. Calculate the Moran's Index, Moran's Scatterplot, and Local Indicator of Spatial Autocorrelation (LISA) Index for spatial autocorrelation analysis;
- d. Create a thematic map that depicts the pattern of COVID-19 spread;
- e. Interpret the results of the analysis; and
- f. Make conclusions based on results.

### 4. Results

#### 4.1. Descriptive analysis

Table 2 reveals that the number of Covid-19 cases, additional COVID-19 cases, and COVID-19 patients under treatment in South Kalimantan Province has been steadily increasing over time, from the beginning of the pandemic to the cases in the new normal era. This pattern is also consistent with the rise in the average case for each district/city in the province.

#### Table 2. Descriptive analysis

| Variable                              | Stage     | Mean     | Minimum | Maximum |
|---------------------------------------|-----------|----------|---------|---------|
| Number of positive COVID-19 cases     | Early Stage | 84.62    | 0       | 452     |
|                                       | PSBB      | 1,131.85 | 76      | 5,568   |
|                                       | New Normal | 5,4661.85 | 16894   | 257,180 |
| Number of additional positive COVID-19 cases | Early Stage | 8.69    | 0       | 32      |
|                                       | PSBB      | 61.92    | 0       | 387     |
|                                       | New Normal | 725.38   | 301     | 2,839   |

| Number of patients COVID-19 in the treatment | Early Stage | 69.85    | 0  | 339 |
|                                              | PSBB       | 876.85   | 25 | 3,876 |
|                                              | New Normal | 18,088.69 | 3,884 | 100,376 |

#### 4.2. Spatial weights matrix

In this research, the Queen Contiguity was used to create the contiguity matrix ($C$). This means that if an area is tangent to each other between its sides and angles with other areas, the weighting is 1 and the other areas are 0.
Figure 3. Contiguity matrix

The following is the standardized spatial weights matrix for South Kalimantan Province:

4.3. Index Moran

The following is the analysis for COVID-19 cases in South Kalimantan Province based on the findings of the Moran Index calculation using equation (1):

Hypotheses

\( H_0 : I = 0 \) (There is no spatial autocorrelation in the number of COVID-19 cases, the number of additional COVID-19 cases and the number of COVID-19 patients under treatment in South Kalimantan Province)

\( H_1 : I \neq 0 \) (There is spatial autocorrelation in the number of COVID-19 cases, the number of additional COVID-19 cases and the number of COVID-19 patients under treatment in South Kalimantan Province)
Table 3. Moran Index

| Variable                      | Stage       | Moran Index (I) | E(I)  | Var(I) | Z_{count} |
|-------------------------------|-------------|----------------|------|--------|-----------|
| Number of COVID-19 cases      | Early Stage | 0.217256       | -0.083333 | 0.022223 | 2.016340^a |
|                               | PSBB        | 0.230795       | -0.083333 | 0.022223 | 2.107159^a |
|                               | New normal  | 0.103846       | -0.083333 | 0.022223 | 1.255590^c |
| Number of additional          | Early Stage | 0.281855       | -0.083333 | 0.022223 | 2.449667^a |
| COVID-19 cases                | PSBB        | 0.156928       | -0.083333 | 0.022223 | 1.611662^b |
|                               | New normal  | 0.036105       | -0.083333 | 0.022223 | 0.801187  |
| Number of COVID-19 patients   | Early Stage | 0.226587       | -0.083333 | 0.022223 | 2.078932^a |
| under treatment               | PSBB        | 0.258702       | -0.083333 | 0.022223 | 2.294358^a |
|                               | New normal  | 0.112178       | -0.083333 | 0.022223 | 1.311481^b |

^a significant at $Z_{\alpha=5\%} = 1.64$
^b significant at $Z_{\alpha=10\%} = 1.28$
^c significant at $Z_{\alpha=15\%} = 1.03$

The results of the Moran Index test (Table 3) show that nearly every $Z_{count}$ value is bigger than $Z_{\alpha}$, so that $H_0$ is rejected or there is a spatial autocorrelation between areas. However in the number of additional COVID-19 cases during the new normal stage there is no spatial autocorrelation because it has a $Z_{count}$ value which is smaller than $Z_{\alpha}$. The resulting distribution pattern is clustered since the difference between the Moran Index value, I, was greater than the expected value, $E(I)$, or $I > E(I)$.

4.4. Moran’s Scatterplot

The Moran's Scatterplots (Figures 5, 6 and 7) reveal that each variable has positive spatial autocorrelation since the majority of the distribution regions are in quadrants I and III. Banjarmasin City and Banjar Regency were known to be in quadrant I (High-High) for all variables, while the regencies of Balangan, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Kotabaru, and Tabalong are known to be in quadrant III (Low-Low).

Figure 5. Moran's scatterplots of the number of COVID-19 cases
Figure 6. Moran's scatterplot of the number of additional COVID-19 cases

Figure 7. Moran's scatterplot of the number of COVID-19 patients under treatment

4.5. Local Indicator of Spatial Autocorrelation (LISA)
The following is the analysis for COVID-19 cases in South Kalimantan Province based on the findings of the LISA Index calculation using equation 8:

Hypotheses

\[ H_0 : I = 0 \]  (At area \( i \) there is no spatial autocorrelation in the number of COVID-19 cases, the number of additional COVID-19 cases and the number of COVID-19 patients under treatment)

\[ H_1 : I \neq 0 \]  (At area \( i \) there is spatial autocorrelation in the number of COVID-19 cases, the number of additional COVID-19 cases and the number of COVID-19 patients under treatment)

Table 4. LISA index

| Region/City | Number of COVID-19 cases | Number of additional positive confirmed COVID-19 cases | Number of COVID-19 patients under treatment |
|-------------|--------------------------|------------------------------------------------------|---------------------------------------------|
|             | Early stage | PSBB | New Normal | Early stage | PSBB | Early stage | PSBB | New Normal |
| Balangan    | 1.055       | 2.646b | 1.737a     | 1.953a      | 1.701a | 1.012       | 1.943a | 1.531      |
| Barito Kuala| 1.540       | 0.517 | -0.153     | 1.014       | 0.111  | 1.528       | 1.221  | -0.021     |
Based on Table 4 for the LISA Index test, it can be seen that Banjar Regency has the values of $Z_{count}$ more than $Z_{a}$, so $H_0$ was rejected for all variables and time periods. Although not in all time periods, Balangan Regency, Hulu Sungai Tengah Regency, Hulu Sungai Utara Regency, Banjarmasin City, Tabalong Regency, and Banjar Regency have the values of $Z_{count}$ greater than $Z_{a}$, so that $H_0$ was rejected, indicating that those regions have the ability to influence the high and low cases in neighboring areas for a particular time period.

4.6. Thematic Map

![Thematic map of the number of COVID-19 cases](image)
Due to the similarity in the number of COVID-19 cases, the number of additional COVID-19 cases, and the number of COVID-19 patients under treatment in neighbouring areas in South Kalimantan Province and based on the thematic maps (Figures 8, 9, and 10), the patterns of regional distributions appear to be clustered. These results are in agreement with the Moran Index analysis, which demonstrated a positive spatial autocorrelation.

5. Discussion

The results of this research were based on a spatial autocorrelation analysis with the Moran and LISA Index tests. It is known that all variables (the number of COVID-19 cases, the number of additional COVID-19 cases, and the number of COVID-19 patients under treatment) have a spatial autocorrelation. However, the Moran Index for the number of additional COVID-19 cases during the new normal is close to zero, implying that there is no spatial autocorrelation and no tight relationship between neighboring areas at that time.

Based on the LISA Index, Banjarmasin City and Banjar Regency have shown a positive spatial autocorrelation in the High-High (HH) category for all variables, except in the beginning of the pandemic and the new normal period. This result implied that the high number of COVID-19 cases in Banjarmasin City and Banjar Regency has an impact on the number of COVID-19 cases in neighboring areas. On the other hand, the regencies of Balangan, Hulu Sungai Tengah, Hulu Sungai Utara, and Tabalong have a spatial autocorrelation in the Low-Low (LL) category for all variables in a given period. This result meant that the low number of Covid-19 cases in those regencies has impact on the low number of COVID-19 cases in their surrounding areas.

Several factors may be responsible for the high number of COVID-19 cases in South Kalimantan Province, including population age [9], population density, population mobility, type of mobility, type of community work, medical history (comorbidities) of COVID-19 patients [2], poor care, and accessibility to health care facilities [10]. This is corroborated by data from the Provincial Bureau
Statistics of South Kalimantan, which shows that Banjarmasin has the highest population density, with 9,134 people/km².

Another situation, such as the low number of COVID-19 cases and the increased number of COVID-19 cases in South Kalimantan Province, is thought to be due to a large number of asymptomatic cases and the community's low ability to conduct COVID-19 testing due to the community's low socioeconomic factors [2]. According to data from the Provincial Bureau Statistics of South Kalimantan, the Hulu Sungai Utara regency and Balangan regency have the greatest percentages of poor people. Furthermore, because PSBB have imposed travel limits to the epicenter area, this is one of the good effects of its deployment in Banjarmasin City, Banjarbaru City, Barito Kuala Regency dan Banjar Regency.

Factors that may have influenced in the number of COVID-19 patients under treatment are socioeconomic status and health facilities (put references here if you have). COVID-19 cases tend to be more common in areas with easy access to medical care and that are cost-conscious. As a result, economically disadvantaged persons may be unable to obtain medical treatment since quality care will be determined by purchasing power [10]. This is in line with the information from the Provincial Bureau Statistics of South Kalimantan (2020) that Banjar Regency is the area with the lowest percentage of poor people and Banjarmasin City is known as the area with the highest number of health facilities (9 general hospitals) so that people can easily access and get medical care for COVID-19. Meanwhile, Hulu Sungai Tengah Regency and the surrounding areas are known to have a high percentage of poor people with only 1 general hospital.

The increasing number of COVID-19 cases in South Kalimantan Province during the normal period that lacked spatial autocorrelation was assumed to be related to regional disparities in mitigation and the state of the region itself. For example, in the northern part of South Kalimantan Province, Banjarbaru City, is suspected to be the cause of an increase in more COVID-19 cases than COVID-19 cases transferred from other regions. Tabalong Regency is known to be the boarder of East Kalimantan Province, which was not included in the study. All variables in all time periods have spatially clustered patterns around the High-High (HH) and Low-Low (LL) categories, with the Low-Low (LL) category is not situated near the epicenter location, indicating that the COVID-19 pandemic is spreading spatially. The findings of this research are consistent with those of studies conducted by [4] and [11]. Although the pattern of distribution is known, the factors that influence the high and low numbers of COVID-19 cases in South Kalimantan Province require further investigation.

6. Conclusion

During the early stages of the COVID-19 pandemic in South Kalimantan Province, Banjarmasin city was recognized to be the epicenter of pandemic until the new normal stage. Based on the Moran Index test, the number of COVID-19 cases and the number of COVID-19 patients under treatment had spatial autocorrelation over the time. Meanwhile, the number of additional positive COVID-19 cases had a spatial autocorrelation only during the pandemic's early stages and PSBB. Based on the LISA Index test, it is known that Banjarmasin City and Banjar Regency have a positive spatial autocorrelation with the High-High (HH) category, while the regencies of Balangan, Hulu Sungai Tengah, Hulu Sungai Utara, and Tabalong have a spatial autocorrelation with Low-Low (LL) category. The distribution pattern for the number of COVID-19 cases, the number of additional COVID-19 cases, and the number of COVID-19 patients under treatment in South Kalimantan Province indicate a cluster pattern, which means that neighboring areas have similar characteristics of COVID-19 cases. Put some recommendations here based on how this research will benefit the government to take actions to reduce the spread of COVID-19.
References

[1] World Health Organization Coronavirus Disease 2019 (COVID-19) in Situation Report-WHO
[2] Cordes J and Castro M C 2020 Spat. Spatiotemporal. Epidemiol. 34 100355
[3] Kementerian Kesehatan RI Situasi Terkini Perkembangan Coronavirus Disease (COVID-19) 30 September 2020 (Jakarta: Kementerian Kesehatan)
[4] Kang D, Choi H, Kim J H and Choi J 2020 Int. J. Infect. Dis. 94 96–102
[5] Adegboye O A, Adekunle A I, Pak A, Gayawan E, Leung D H, Rojas D P, F. E, S. M E and Eisen D P 2021 Travel Med. Infect. Dis. 40 101988
[6] Anselin L 1993 Exploratory Spatial Data Analysis and Geographic Information Systems National Center for Geographic Information and Analysis of California (Santa Barbar: CA93106)
[7] Lee J and Wong D W 2001 Statistical analysis with ArcView GIS (New York: John Wiley & Sons)
[8] Kosfeld R, Eckey H F and Türck M 2007 WiSt-Wirtschaftswissenschaftliches Stud. 36 157–162
[9] Bermudi P M M, Lorenz C, Aguiar B S, Failla M A, Barrozo L V and Chiaravalloti-Neto F 2021 Travel Med. Infect. Dis. 39 101945
[10] Iyanda A E, Adeleke R, Lu Y, Osayomi T, Adaralegbe A, Lasode M, Chima-Adaralegbe N J and Osundina A M 2020 A J. Infect. Public Health 13 1438–1445
[11] Kim S and Castro M C 2020 Int. J. Infect. Dis. 98 328–333