Analysis of spatial autocorrelation of diarrhea events in East Kalimantan in 2015-2017 using a local indicator of spatial autocorrelation

M N Hayati*, I Purnamasari, and S Wahyuningsih
Department of Statistics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia

*Corresponding author: meminorhayati@yahoo.com

Abstract. Diarrhea is still a public health issue in developing countries, one of which is Indonesia. This is indicated by an increase in diarrhea morbidity rates from year to year. Therefore, it was necessary to conduct research related to diarrheal diseases. Factors that influence the incidence of diarrhea in an area depend on the condition of the environment or the surrounding area. Spatial autocorrelation is the correlation of a variable with itself based on space or can be said similarity of objects in space, both in terms of distance, time or region. Local Indicator of Spatial Autocorrelation (LISA) is a method for measuring and identifying spatial autocorrelation locally. The purpose of this study was to determine whether there is spatial autocorrelation in the incidence data of diarrhea in East Kalimantan in 2015-2017 by using LISA testing and to determine the level of vulnerability to the incidence of diarrhea in East Kalimantan. The results of data analysis with $\alpha = 10\%$ indicated that there were spatial autocorrelations in several districts/cities in East Kalimantan in 2015, 2016 and 2017. The results of the analysis also showed that Samarinda City and Mahakam Ulu District were categorized as prone to diarrhea.

1. Introduction
The emergence of diseases caused by places that are less environmentally friendly is a problem for developing countries. Until now diarrhea is still a public health problem in developing countries, one of them is in Indonesia. This is due to the increasing number of diarrhea morbidity from year to year. Based on the pattern of causes of death of all ages, diarrhea is the cause of death ranked 13th with a proportion of 3.5%. While based on infectious diseases, diarrhea is the third leading cause of death after TB and Pneumonia [1].

In 2015 the number of diarrhea cases in East Kalimantan reached 71,104. In 2016 the number of diarrhea cases in East Kalimantan reached 76,482. In the following year, 2017, the number of diarrhea cases in East Kalimantan reached 76,511 [2], which means there was an increase in the number of diarrhea incidents in the last three years.

Seeing the high incidence of diarrhea, it is necessary to do research related to the disease. Factors that influence the incidence of diarrhea in a particular region also depend on the condition of the environment or the area around it. This is certainly related to Tobler's Law (Tobler's first law of geography) which states that "everything is related to everything else, but things are more related than distant things", everything is related to one another, but something closer will be more influential than
something far away [3]. It can be concluded that the correlation (relationship) in a space is a natural thing that happens between one location with another location, causing spatial autocorrelation.

Spatial autocorrelation is the correlation between variables with themselves based on space or it can be said that the similarity of objects in space, both distance, time or region. The identification of patterns of spatial autocorrelation locally is an important problem, meaning finding spatial correlation at each location. For this purpose, the Local Indicator of Spatial Autocorrelation (LISA) can be used to identify how the relationship between a location of observation and other observation locations is related. The higher the LISA Index, provides information that adjacent areas have almost the same value or form a clustered spread [3].

Based on this, it is necessary to do research related to the analysis of spatial autocorrelation of diarrhea events in East Kalimantan in 2015-2017 using a local indicator of spatial autocorrelation.

2. Materials and Methods

2.1. Data Source
The variables used in this study were the percentage of the incidence of diarrhea in East Kalimantan in 2015-2017 by district/city.

2.2. Analysis Stage
The data analysis technique used in this study is spatial analysis using LISA. Data analysis in this study was carried out by the following steps:

a. Making descriptive statistics on the percentage of diarrhea in East Kalimantan in 2015-2017.

b. Determine the closeness between districts/cities in each region in East Kalimantan by making a contiguity matrix. For determining the closeness between districts/cities, the tool used is a map of East Kalimantan and the method used in determining proximity using queen contiguity. East Kalimantan consists of 10 districts/cities, then a 10 x 10 contiguity matrix is formed.

c. Calculating spatial weighting matrices

d. Look for the LISA Index value

e. Testing the parameters \( I_i \) and the hypothesis are:

\[
H_0 : I_i = 0, \ i=1,2,...,10 \\
H_1 : I_i \neq 0, \ i=1,2,...,10
\]

f. Make a vulnerability map for diarrhea events from 10 districts/cities in East Kalimantan using the LISA test results.

g. Conclude the results of the analysis

2.3. Spatial pattern
The spatial pattern is something that is related to the placement or arrangement of objects on the surface of the earth. Analysis of spatial patterns from point distribution by distinguishing locations and attributes can be done using analysis of spatial autocorrelation [3].

2.4. Spatial Autocorrelation
Correlation of a variable with itself based on space or a measure of the similarity of objects in space (distance, time and region) is referred to as spatial autocorrelation. If there is a systematic pattern in the distribution of a variable, it can be said that there is a spatial autocorrelation. Spatial autocorrelation indicates that attribute values in certain regions are related to these attribute values in other regions that are adjacent or neighboring.

Spatial autocorrelation at a point that is the extent to which a point is interconnected or things that occur at these points are similar to other points. Positive spatial autocorrelation indicates the
significance of the distribution points having the same characteristics, which tend to be close to each other. Negative spatial autocorrelation shows that adjacent distribution points tend to have different characteristics. If the point is randomly distributed then there is no spatial autocorrelation. Visually, can be seen in figures 1 to figure 3 [3].

The determination of the proximity relationship between locations can be expressed in a contiguity matrix or spatial weighting matrix.

2.5. Contiguity Matrix
Matrix contiguity is a matrix that describes the relationship between regions. The contiguity matrix will give a value of 1 to the area directly adjacent to the location of observation and the rest given 0 or left blank. The contiguity matrix shows the spatial relationship of an area to other neighboring regions. Granting of value 1 is given if region-i is directly adjacent to region-j while a value of 0 is given if region -i is not adjacent to region-j. The contiguity matrix is often called the binary matrix and also connectivity matrix, which is denoted by:

\[ c_{ij} = \sum_{j=1}^{n} c_{ij} \tag{1} \]

with:

- \( c_{li} \) = total i line element
- \( c_{ij} \) = value of the i-row element of the j-column.

The value of \( c_{ij} \) is 1 if the region-i is adjacent to the region-j and \( c_{ij} \) is 0 if the region-i is not neighboring to the j-region.

The contiguity matrix has a public grid of neighborhoods that can be defined in [3]: side contact (rook contiguity), corner intersection (Bhisop contiguity) and corner side contact (Queen contiguity).

2.6. Spatial Weighting Matrix
Spatial weighting matrix is a matrix that is symmetrical and has a main diagonal which is always zero [3]. Giving a weighting code in addition to binary code can also be made in the form of Row Standardization. Row Standardization is based on the number of neighbors on the same line in the contiguity matrix. The formulas from Row Standardization are as follows:

\[ w_{ij} = \frac{c_{ij}}{c_{li}} \tag{2} \]
2.7. Local Indicator of Spatial Autocorrelation (LISA)

LISA can be used to identify local autocorrelation coefficients or spatial correlations at each location. LISA is a new technique of analyzing exploration data for visualization needs. LISA here identifies how the relationship between a location of observation and the location of another observation. The index is as follows [4]:

\[
I_i = \frac{1}{m_2} \sum_{j \neq i} w_{ij} z_j
\]

(3)

with:

\[z_i = (x_i - \bar{x})\]

\[z_j = (x_j - \bar{x})\]

\[m_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}\]

Testing of the \(I_i\) parameter can be done by:

\(H_0: I_i = 0\) (there is no spatial autocorrelation/there is no relationship between the i-location and the location adjacent to it)

\(H_1: I_i \neq 0\) (there is spatial autocorrelation/there is a relationship between the i-location and the location adjacent to it)

Test statistics:

\[
Z(i) = \frac{I_i - E(I_i)}{\sqrt{\text{var}(I_i)}}
\]

(4)

\(I_i\) is the LISA index, \(Z(i)\) is the value of the LISA index test, \(E (I_i)\) is the expected value of the LISA index, \(\text{var}(I_i)\) is the variance value of the LISA index.

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

(5)

\[
\text{var}(I_i) = w^{(2)}_i \left( \frac{n - m_1}{m_2} \right) + 2w_{2i} \left( \frac{2\sum_{i,j} w_{ij}^2 - n}{n-1(n-2)} \right) - \frac{w_i^2}{(n-1)^2}
\]

(6)

with:

\[
w^{(2)}_i = \sum_{j \neq i} w_{ij}^2, i \neq j
\]

\[w_i = \sum_{j \neq i} w_{ij}, i \neq j\]

\[2w_{2i} = \sum_{k=1}^{n} \sum_{k \neq i} w_{ik} w_{ij}\]

\[m_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{n-1}\]

This test will reject the null hypothesis if \(|Z(i)| > Z(\alpha/2)\) or p-value < \(\alpha\).
2.8. Diarrhea
Diarrhea is a disease that occurs when there is a change in the consistency of stool with a frequency of bowel movements three or more times. Diarrhea is an infectious disease that is a public health problem and if the treatment is carried out too late it will cause death [5].

2.9. Transmission of Diarrhea
Various disease agents generally ride on air, water, food, insects or human media through direct contact. Various agents of disease and their media are referred to as environmental components that have potential disease hazards. Environmental components that have the potential to cause various kinds of diseases include water, food, insects, air and humans. Diarrhea is an environmental-based disease, with a chain of transmission through the medium of water, food, insects, and humans [6].

3. Results and Discussion

3.1. Descriptive statistics
The following are statistical results description of the percentage of diarrhea in East Kalimantan in 2015-2017 in 10 districts/cities namely Paser District, West Kutai District, Kutai Kartanegara District, East Kutai District, Berau District, North Penajam Paser District, Balikpapan City, Samarinda City, Bontang City, and Mahakam Ulu District.

| Year | Minimum | Maximum | Mean |
|------|---------|---------|------|
| 2015 | 1.422   | 5.676   | 2.575|
| 2016 | 1.456   | 4.209   | 2.625|
| 2017 | 1.426   | 4.592   | 2.591|

Based on the results in Table 1, it is known that the average percentage of the highest incidence of diarrhea in East Kalimantan is 2.625% in 2016. The lowest percentage of incidence of diarrhea is in East Kalimantan which is 2.575% in 2015.

3.2. LISA Testing
After doing descriptive statistical analysis, then the spatial autocorrelation test is carried out locally with LISA to see which locations in 2015-2017 have spatial autocorrelation using R software. Hypothesis testing locally with the significance level used is $\alpha = 10\%$. Examination of spatial autocorrelation locally is presented as follows:

$H_0 : I_i = 0, i=1,2,...,10$ (there is no spatial autocorrelation/there is no relationship to the district/city $i$ in East Kalimantan with the district/city that is located adjacent)

$H_1 : I_i \neq 0, i=1,2,...,10$ (there is spatial autocorrelation/there is a relationship to the district/city $i$ in East Kalimantan with the district/city that is located adjacent)

This test will reject the null hypothesis if p-value $<\alpha=10\%$.
Table 2. P-value for 2015-2017

| I | District/City             | Year |        |        |        |
|---|---------------------------|------|--------|--------|--------|
|   |                           | 2015 | 2016  | 2017  |
| 1 | Balikpapan                | 0.853| 0.899  | 0.963 |
| 2 | Berau                     | 0.845| 0.905  | 0.820 |
| 3 | Bontang                   | 0.359| 0.215  | 0.225 |
| 4 | West Kutai                | 0.703| 0.813  | 0.843 |
| 5 | Kutai Kartanegara         | 0.980| 0.846  | 0.546 |
| 6 | East Kutai                | 0.706| 0.489  | 0.731 |
| 7 | Mahakam Ulu               | 0.020*| 0.041*| 0.041*|
| 8 | Paser                     | 0.703| 0.478  | 0.850 |
| 9 | North Penajam Paser       | 0.799| 0.904  | 0.844 |
| 10| Samarinda                 | 0.344| 0.024*| 0.097*|

*Significant at $\alpha = 10\%$

The results in Table 2 show that in 2015 there was a spatial autocorrelation of the incidence of diarrhea in Mahakam Ulu District with districts/cities located adjacent to Mahakam Ulu District because of the p-value (0.020) < $\alpha = 0.1$. In 2016 and 2017 there is a spatial autocorrelation of the incidence of diarrhea in Samarinda City and Mahakam Ulu District with districts/cities that are located adjacent to the two districts/cities because the p-value < $\alpha = 0.1$. The significant map of spatial autocorrelation of diarrhea events in East Kalimantan is shown in figure 4 and figure 5 below:

![Figure 4. Significant maps of 2015 using Arcview GIS software](image.png)
3.3. Map of potential areas of potential diarrhea

Map of vulnerability of potentially diarrhea areas in East Kalimantan is used to prioritize areas that need to be considered to reduce the incidence of diarrhea.

The vulnerability map in figure 6 is made based on the results of the LISA test for 2015-2017 which are divided into 2 categories, namely the safe and vulnerable categories. Samarinda City and Mahakam Ulu District fall into the category of prone to diarrhea because during the year 2015-2017 they experience successive spatial autocorrelations. Paser District, West Kutai District, Kutai
Kartanegara District, East Kutai District, Berau District, North Penajam Paser District, Balikpapan City and Bontang City fall into the safe category of diarrhea because in 2015-2017 there was no spatial autocorrelation.

4. Conclusions
The results obtained from the testing of spatial autocorrelation locally using LISA showed that in 2015 the regions with spatial autocorrelation were Mahakam Ulu District. In 2016 and 2017 regions that have spatial autocorrelation are Mahakam Ulu District and Samarinda City. Regions that are categorized as vulnerable and need to be considered to reduce the incidence of diarrhea are Mahakam Ulu District and Samarinda City.

Acknowledgment
Thanks are extended to the Faculty of Mathematics and Natural Sciences of Mulawarman University who helped in funding this research.

References
[1] Kemenkes RI 2011 Situasi Diare di Indonesia (Jakarta: Kementerian Kesehatan RI) p 3
[2] BPS 2018 Kalimantan Timur Dalam Angka 2018 (Samarinda: Badan Pusat Statistik Provinsi Kalimantan Timur) p 146
[3] Jay L and David W S W 2001 Statistical Analysis with ArcviewGIS (New York: John Wiley & Sons, Inc.)
[4] Luc A 1995 Local Indicator of Spatial Association-LISA J. Geographical Analysis 27 98-100
[5] Widoyono 2008 Penyakit Tropis Epidemiologi, Penularan, Pencegahan dan Pemberantasannya (Surabaya: Erlangga)
[6] Umar F A 2011 Dasar-dasar Penyakit Berbasis Lingkungan (Jakarta: Rajawali Pers)