Spatiotemporal Dengue Disease Clustering by Means Local Spatiotemporal Moran’s Index

I G N M Jaya¹,², Y Andriyana², B Tantular³, Zulhanif⁴, BN Ruchjana⁵

¹,²,³,⁴ Statistics Department, Padjadjaran University, Bandung, Indonesia
⁵ Mathematics Department, Padjadjaran University, Bandung, Indonesia
⁶ Faculty of Spatial Science-Groningen University, Netherlands

mindra@unpad.ac.id

Abstract. Spatiotemporal analysis has been used widely to explain some geographic phenomenon, especially in an epidemiology study. Spatial and temporal autocorrelation coefficients are usually used to assess the spatial and temporal dependencies in set geographic events. However, those statistics are usually computed separately and may lead to the misleading conclusion. Analysing spatiotemporal autocorrelation would be useful to understand the geographical evolution simultaneously. Spatiotemporal autocorrelation can be used to identify the spatiotemporal clustering and outlier via local spatiotemporal autocorrelation. This paper develops a method to estimate and test the local spatiotemporal autocorrelation based on the local spatial Moran’s Index. Randomization permutation test is used to obtain the p-value which is used to construct the disease clustering. The method was applied to identify the spatiotemporal clustering and outlier detection for dengue disease data in Bandung city. Based on image analysis, this method presents the better result compare than the local spatial Moran’s Index which is done for every time separately.

1. Introduction
Dengue disease is an injurious infectious disease transmitted via the female Aedes-spp. mosquito from one person to another [1]. Dengue disease is globally endemic in most (sub) tropical countries, and particularly in urban and semi-urban regions. The disease appears to be continuing to spread rapidly. The transmission of pathogens of dengue disease is promoted by multiple risk factors such as environmental and climates factors. Analysing the spatiotemporal pattern of dengue disease may be useful to develop the etiological hypothesis of the disease with regard to environmental or lifestyle-related risk factors [2]. Where the etiology of a disease is not yet well-established, e.g., dengue disease, it may be important to determine whether the cases of the disease are occurring independently or if they seem to be related [3]. In this research, we will be concerned with the condition in which related disease cases occur closely together, both in space and time and probability create the spatiotemporal clusters. Statistical methods have been widely used for spatial cluster identification in epidemiology study, especially for the processing of geographic data. Those methods are: join-count statistics [4]; Ohno statistics [5]; Poisson statistics [6]; Global Moran’s I ( [7]; [8]); Global Geary’s C ( [7]; [8]); General Getis-Ord’s G [9]; Local Moran’s I ; and Local $G_i(d)$ and $G_i^*(d)$ ([9]; [10]; [11]). Statistical approach
for spatial autocorrelation such as the Moran’s I and Geary’s C methods present a global autocorrelation for overall degree of spatial autocorrelation in a dataset. However, all of the methods only accommodate the spatial variation which is ignoring the temporal attribute. Ignoring temporal attribute may lead to the misleading conclusion. In this paper, we develop Local spatiotemporal Moran’s Index which accommodates the temporal dependencies. The p-value of local spatiotemporal Moran’s Index is used to define disease clustering which developed based on randomization permutation test.

The structure of the remainder of this paper is as follows; Section 2 presents the spatiotemporal Moran’s Index, Section 3 applies the randomization method to obtain p-value. Section 4 present application for dengue disease data in Bandung City and section 5 presents the discussion.

2. Spatial and spatiotemporal clustering

2.1. Spatial Clustering

Spatial clustering can be seen as a spatial autocorrelation function. The high spatial autocorrelation indicates there is a spatial cluster in the data set. There are two measurement of spatial autocorrelation. Both are global and local spatial autocorrelation. Global and local Moran’s Index are commonly used in geographical study to identify the global and local dependencies for a set geographical data. Global autocorrelation measure the spatial dependencies in global way where the autocorrelation coefficient is unique for all regions. The coefficient of global Moran’s Index value between -1 and 1. The value closed to positive one indicate there is a spatial cluster. Global Moran’s index is based on regression coefficient the observed variables with its lag dependence. Global Moran’s I can be defined as [12]; [10]:

\[
I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} (y_i - \bar{y})^2}
\]  

(1)

where \(n\) denotes the number of spatial units being analysed; \(y_i\) is the attribute value of spatial object \(i\); \(\bar{y}\) is the mean of all \(y\). Spatial weight matrix \(w_{ij}\) denotes the element of spatial weight matrix for observation of spatial unit \(i\) and spatial unit \(j\):

\[
w_{ij} = \begin{cases} 
1 & \text{if } i \text{ and } j \text{ are connected} \\
0 & \text{otherwise}
\end{cases}
\]  

(2)

The second, is local Moran’s Index. This is commonly used to detect the outlier or hotspot in geographical data set. This method is useful for identifying high risk cluster region. The value of local Moran’s does not have to be interval -1 to 1. The larger positive value of local Moran’s Index indicates there is a strong local spatial cluster. For spatial Moran’s Index can be defined as:

\[
I_i = \frac{n(y_i - \bar{y}) \sum_{j=1}^{n} w_{ij} (y_j - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]  

(3)

The normal approximation (Z-test) and permutation test are usually used to test the signification of Global and Local Moran’s I.

2.2. Spatiotemporal clustering

Spatial Moran’s Index can be extended becomes spatiotemporal Moran’s Index to accommodate the temporal autocorrelation. Temporal autocorrelation is an important component in evaluating the spatial evolution of the geographical phenomenon. Ignoring temporal attributes may lead to the misleading conclusion because of the number of incidence in location \(i\) at time \(t\) are connected not only with location \(j\) but also time \(s\). The local spatiotemporal autocorrelation can be used to identify clustering of regions and time periods (where and when) simultaneously (spatiotemporal Hot Spot). Here, we develop spatiotemporal Moran’s Index based on Autoregressive model order one (\(AR 1\)). We assume, the observed variable \(y_{it}\) are related to the \(y_{it-1}\) and \(y_{it+1}\) in another word, \(w_{st} = 1 \text{ if } |s - t| < 1\). Using the Spatiotemporal Moran’s Index, it was possible to calculate Global and Localized indices that account for spatial and temporal autocorrelation simultaneously.
The spatiotemporal weight matrix have to be measured in more general way. We only need to restructured the spatial weight matrix (2) becomes spatiotemporal weight matrix. Spatiotemporal weight have to account for the spatiotemporal autocorrelation between $y_{it}$ and $y_{js}$, defined as [13]:

$$
\tilde{w}_{(it,js)} = \begin{cases} 
    w_{ij}, & \text{if } t = s \\
    1, & \text{if } i = j \text{ and } |t - s| = 1 \\
    0, & \text{otherwise}
\end{cases}
$$

(4)

where $w_{ij}$ is defined in (2). Hence, the spatiotemporal Global Moran’s Index (MoranST) is defined as [13]:

$$
\text{MoranST} = \frac{nT \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j=1}^{n} \sum_{s=1}^{T} \tilde{w}_{(it,js)} (y_{it} - \bar{y}) (y_{js} - \bar{y})}{\sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j=1}^{n} \sum_{s=1}^{T} \tilde{w}_{(it,js)} \sum_{s=1}^{T} (y_{it} - \bar{y})^2}
$$

(5)

where $\bar{y}$ is the mean of the observed residuals $r_{it}$ over $T$. MoranST coefficient between -1 to 1. The value close to 1 corresponds strong positive spatiotemporal autocorrelation, while 0 corresponds to complete spatiotemporal randomness.

For spatiotemporal Local Moran’s index is developed based logical thinking of spatial local Moran’s Index with substitute the spatial weight matrix by spatiotemporal weight matrix in (4). From (3) the spatial local Moran’s index can be written as:

$$
\text{MoranST}_{it} = \frac{n(y_{it} - \bar{y}) \sum_{j=1}^{n} \sum_{s=1}^{T} \tilde{w}_{(it,js)} (y_{js} - \bar{y})}{\sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \bar{y})^2}
$$

(6)

The value of Local spatiotemporal local Moran’s I does not have to be interval -1 to 1.

The complexity in developing the spatiotemporal Moran’s Index is how to find out the exact statistical testing for it. Spatiotemporal Moran’s Index includes spatial and temporal variation which have to accommodate in developing statistical testing. [14] uses simple technique which assumes the statistical test for spatiotemporal Moran’s Index is similar with spatial Moran’s Index with took the strong assumption that for event $i$ and event $j$, the spatiotemporal weight is defined as $w_{ij} t_{ij}$. It means, the spatial and temporal component have a similar event.

Spatial weight matrix $w_{ij}$ denotes the element of spatial weight matrix for observation of spatial unit $i$ and spatial unit $j$;

$$
w_{ij} = \begin{cases} 
    1 & \text{if } i \text{ and } j \text{ are connected} \\
    0 & \text{otherwise}
\end{cases}
$$

Temporal weight matrix $t_{ij}$ denote the element of temporal weight matrix for observation of time unit $i$ and time unit $j$;

$$
t_{ij} = \begin{cases} 
    1 & \text{if } |t_i - t_j| \leq t_0 \\
    0 & \text{otherwise}
\end{cases}
$$

(7)

It means, the temporal index $(i, j)$ have to follow spatial index $(i, j)$ where this assumption is not realistic because we cannot calculate a spatiotemporal weight for $w_{ij} t_{kl}$. Using [14] assumptions, we cannot take into account the pair of observations $y_{it}$ and $y_{js}$, we only use observations $y_{ii}$ and $y_{jj}$. It means we only take into account observation at location $i$ and time $i$ and observation at location $j$ time $j$. Here we developed a permutation test of spatiotemporal autocorrelation.

2.3. Random Permutation Test of Spatiotemporal Autocorrelation

Random permutation test becomes a popular testing procedure in hypothesis testing because it does not rely need assumption about the distribution of the data. In geographical analysis, the normality assumption is hard to be satisfied. The basic idea of the random permutation test is a central limit theorem. The $p$-value is used in testing obtained by resampling on the observed data form many times, where $p$-value presents the probability of obtaining data as extreme as the observed data when the null hypothesis is true [15]. If the data are drawn many times with replacement in case the null hypothesis being true, the number of cases with data as extreme as the observed data could be counted, and a $p$-
value calculated. The steps of the random permutation test for spatiotemporal autocorrelation are as follow [15] :
1. Define spatiotemporal weight matrix as in (4)
2. Let \( y^0 = (y_{11}, y_{12}, ..., y_{nT}) \) denotes the vector observed spatiotemporal data and construct the corresponding value of spatiotemporal Global or Local Moran’s I, \( f^0 = f(y^0) \) where \( f \) can be a function of Global and Local spatiotemporal Moran’s I
3. Simulate \( M \) random permutations, \( p^j = (p^j_1, ..., p^j_{nT}) \) of integer \((1,2,3, ..., nT)\)
4. For each permutation, \( p^j \), construct the corresponding permuted data vector, \( y(p^j) = (y_{p^j_1}, ..., y_{p^j_{nT}}) \) and resulting \( f \), denote by \( f^j = f[y(p^j)], j = 1,2, ..., M \)
5. Rank the values \( f^0, f^1, ..., f^M \) from low to high, so that if \( f^j \) is the \( k^{th} \) highest value then \( rank(f^j) = k \)
6. If \( rank(R^0) = k \) then construct p-value for this test to be
\[
\alpha = \frac{k}{M + 1}
\]
7. Decision for testing
(i) If \( \alpha \) is low (e.g., \( \alpha \leq 0.05 \) ) indicates there is significant positive spatiotemporal autocorrelation.
(ii) Inversely, if \( \alpha \) is high (e.g., \( \alpha \geq 0.95 \) ) indicates there is significantly negative spatiotemporal autocorrelation
(iii) If neither (i) or (ii) satisfied, indicate the spatiotemporal independence hypothesis, \( H_0 \), cannot be rejected.

3. Application: Dengue Disease in Bandung City
Bandung is the most economically advanced city in West Java. The number of dengue incidence usually high for every year. In this section, to illustrate the Moran’s Index for spatiotemporal data, dengue data in 30 Bandung regions over the period January-December, 2016 have been used. In 2016 there is 3880 number of dengue cases in Bandung. Coblong, Buah Batu, and Rancasari are the three sub-districts with the highest number of cases. The number of incidences is increasing from January to May, and the highest dengue incidence in May. Boxplot present there are some spatial and time have the highest incidence which can be diagnostic as the outlier.

![Boxplot of Number of Incidence by Sub District](a) Boxplot of Number of Incidence by Sub District

![Boxplot of Number of Incidence by Sub District](b) Boxplot of Number of Incidence by Sub District

**Figure 1.** Distribution of Dengue disease in Bandung, 2016
Spatiotemporal distribution of dengue disease shows that the number incidence of dengue disease is different for every location and time. Figure 2(b) present there is a temporal autocorrelation in the data which are several time lags are significant.

**Figure 2.** (a) Spatiotemporal Distribution of Dengue disease in Bandung, 2016 and (b) Partial Spatiotemporal Autocorrelation Function of dengue incidence

| Month | Moran's I | p-value |
|-------|-----------|---------|
| Jan   | 0.303     | 0.005** |
| Feb   | 0.095     | 0.115   |
| Mar   | 0.301     | 0.005** |
| Apr   | 0.385     | 0.002** |
| May   | 0.295     | 0.003** |
| Jun   | 0.319     | 0.005** |
| Jul   | 0.072     | 0.175   |
| Aug   | 0.219     | 0.018*  |
| Sep   | 0.001     | 0.358   |
| Oct   | 0.300     | 0.005** |
| Nov   | 0.183     | 0.037*  |
| Dec   | 0.236     | 0.018** |

Note:
** significant at level $\alpha < 0.01$
* significant at level $0.01 < \alpha < 0.05$
| District               | Moran's I | p-value  |
|-----------------------|-----------|----------|
| Gedebage              | 0.659     | 0.005**  |
| Ujungberung           | 0.737     | 0.001**  |
| Cinambo               | 0.441     | 0.006**  |
| Bandung Kulon         | 0.451     | 0.039*   |
| Andir                 | 0.453     | 0.034*   |
| Babakan Ciparay       | 0.406     | 0.044*   |
| Bojongloa Kaler       | 0.385     | 0.046*   |
| Sukajadi              | 0.622     | 0.004**  |
| Cidadap               | 0.079     | 0.291    |
| Coblong               | 0.604     | 0.010*   |
| Cicendo               | 0.298     | 0.084*   |
| Bandung Wetan         | 0.205     | 0.145    |
| Sumur Bandung         | 0.307     | 0.104    |
| Batununggal           | 0.628     | 0.004**  |
| Cibeunying Kidul      | 0.383     | 0.038*   |
| Regol                 | 0.443     | 0.022*   |
| Bandung Kidul         | 0.404     | 0.027*   |
| Astanaanyar           | 0.292     | 0.102    |
| Kiaraccondong         | 0.719     | 0.001**  |
| Buah Batu             | 0.472     | 0.020*   |
| Mandala Jati          | 0.591     | 0.008**  |
| Cibiru                | 0.399     | 0.024*   |
| Bojongloa Kidul       | 0.268     | 0.104    |
| Cibeunying Kaler       | 0.484     | 0.032*   |
| Panyileukan           | 0.526     | 0.018*   |
| Antapani              | 0.580     | 0.008**  |
| Lengkong              | 0.459     | 0.027*   |
| Sukasari              | 0.480     | 0.026*   |
| Rancasari             | 0.639     | 0.005**  |
| Arcamanik             | 0.604     | 0.002**  |

Table 1 presents for several months the spatial Moran’s Index are significant at level 0.05. Table 2 presents the temporal Moran’s Index for 30 sub-districts in Bandung. Several sub-districts have high values of Moran’s Index. We also compute the Local spatial Moran’s Index by month as presented in Figure 3.
Figure 3. Local Spatial Moran’s Index for Dengue Disease in Bandung

Figure 3 shows the local spatial Moran’s index which is computing separately by month. However, the Moran’s statistics fail to identify temporal cluster and the outlier in Coblong and Buah Batu sub-districts. The figure could be present the clear spatiotemporal clustering and outlier. This is because of ignoring the temporal dependencies. From that information it is important to estimate the spatiotemporal Moran’s Index simultaneously.

Figure 4. Spatiotemporal Moran’s Index with Permutation Test

The dengue incidence data in Bandung city has high spatiotemporal Moran’s Index coefficient, 0.6113 with small p-value, 0.001 (see Figure 4). It means there is a spatiotemporal dependency in the data. The observation in location \( i \) and time \( t \) is related to the neighbouring location at time \( t - 1 \). We also presented the Spatiotemporal Local Moran’s Index in isopleth maps. The map present the spatiotemporal cluster with high incidence and the outlier detection.
Figure 5. Spatiotemporal Moran’s Index for dengue disease in Bandung city

The districts with high incidence are Coblong, Buah Batu and Rancasari. Figure 5(a), Coblong is an outlier because there only one district in north of Bandung has high incidence. However, Buah batu and Rancasari create a cluster for high incidence district at May and June. This result present that spatiotemporal Moran’s Index is a valuable tool to identify the spatiotemporal cluster and outlier.
We compare the spatial Moran’s index which is computed by months separately and spatiotemporal Moran’s Index. We can see there is a significant difference between both statistics. Spatiotemporal Moran’s Index presents clearly spatiotemporal cluster and outlier. Here we present in detail the spatiotemporal clustering based on number of dengue cases in Bandung city.

(c) Image Analysis of Matrix Spatial Moran’s Index

(d) Image Analysis of Matrix p-value Spatial Moran’s Index

Figure 6. Comparison of Spatial Moran’s Index and Spatiotemporal Moran’s Index
Figure 7. Spatiotemporal clustering of dengue disease in Bandung city, 2016 by means local spatiotemporal Moran’s I

Figure 7 displays Spatiotemporal clustering of dengue disease in Bandung city, 2016 by means local spatiotemporal Moran’s I. According to the distribution of the number of cases pattern in Figure 2 and spatiotemporal autocorrelation in Figure 5, using permutation test we obtain the spatiotemporal clustering as a group of significant spatiotemporal autocorrelation. There are five different labels (1) high-high (2) low-low (3) high-low (4) low-high and (5) not significant. Our method presents clearly that the high-risk region with large number incidence found in January up to July. Those months have a number of incidences larger than the other month. The cluster of the high-risk region change over time. This result indicates that there is temporal characteristics that change the clustering structure by time for example there a change in environmental factors and weather. This information may be very useful information for the health department in an effort to overcome dengue disease problem in Bandung city through understanding the distribution patterns of disease diseases - where and when high case rates will be found.

4. Discussion
Temporal autocorrelation is important component in spatiotemporal analysis of geographical data set. Ignoring temporal component might be lead to the misleading result. In order to identify the spatiotemporal cluster of dengue disease data, we develop a local spatiotemporal autocorrelation where the cluster can be defined as grouping regions with significant local spatiotemporal autocorrelation. Here we used permutation test to test the significant of global and local spatiotemporal Moran’s I. We developed R package which can be used to compute Global and Local spatial, temporal and spatiotemporal Moran’s I for Autoregressive Order 1 model and can be developed for high order of AR. We compare the spatial and Spatiotemporal Moran’s Index which produce very different results. Image analysis shows clearly that the spatial Moran’s Index for each time, and spatiotemporal Moran’s Index present different cluster and outlier. The spatial Moran’s Index for each time periods produces under estimate spatiotemporal autocorrelation compare than spatiotemporal Moran’s Index which computed
simultaneously. The spatial Moran’s Index also fails to detect the outlier. As we see from the original data, sub-district Coblong has highest number incidence of dengue disease which is different from the others sub-districts. For further research, the relative risk estimate could be combined with local spatiotemporal Moran’s Index to construct map with clear pattern of geographical disease evolution and disease clustering. The complete approach of disease mapping can be seen in [16].

Acknowledgement
This paper is funded by RFU Unpad contract: 1732 d/UN6.RKT/LT/2018. The authors thank Rector Universitas Padjadjaran and to the anonymous referee whose valuable checking has improved this paper.

References
[1] Acharya K B, Cao C, Lakes T, Chen W and Naeem S 2016 *BMC Public Health* **16** 1
[2] Lione G and Gonthier P 2016 *Analytical and Theoretical Plant Pathology* **106** 19
[3] Mantel N 1967 *Cancer Res* **27** 209
[4] Upton G J and Fingleton B 1985 *Spatial data analysis by example* vol 1 (United Kingdom: John Wiley & Sons Ltd)
[5] Ohno Y, Aoki K and Aoki N 1979 *Int J Epidemiol* **8** 273
[6] Miyawaki N and Chen S C 1981 *Social Science & Medicine Part D: Medical Geography* **15** 93
[7] Cressie N 1993 *Statistics for Spatial Data* (New York: Wiley-Interscience)
[8] Legendre P and Legendre L 1998 *Numerical Ecology* (Amsterdam: Elsevier)
[9] Getis A and Ord J 1992 *Geographical Analysis* **24** 189
[10] Anselin L 1995 *Geographical Analysis* **27** 93
[11] Ord J and Getis A 1995 *Geographical Analysis* **27** 286
[12] Chen Y 2013 *PLOS ONE* **8** 1
[13] Anderson C and Ryan L M 2017 *Int J Environ Res Public Health*. **14** 1
[14] Lee J and Li S 2016 *Geographical Analysis* **1** 36
[15] Good P 2005 *Permutation, Parametric and Bootstrap Tests of Hypotheses* (USA: Springer)
[16] Jaya I G N M, Folmer H, Ruchjana B N, Andriaya Y and Kristianti F 2017 *Regional research frontiers* vol 2 (USA: Springer)