

PAPER • OPEN ACCESS

## Detecting Spatial Autocorrelation for a Small Number of Areas: a practical example

To cite this article: Aswi Aswi *et al* 2021 *J. Phys.: Conf. Ser.* **1899** 012098

View the [article online](#) for updates and enhancements.



The banner features a decorative top border with a repeating pattern of red, white, and blue diagonal stripes. On the left, the ECS logo is displayed in green and blue, followed by the text 'The Electrochemical Society' and 'Advancing solid state & electrochemical science & technology'. To the right of this text is a logo for the 18th International Meeting of the Solid State Ionics Society (IMCS18). The main text of the banner reads '239th ECS Meeting with IMCS18', 'DIGITAL MEETING • May 30-June 3, 2021', and 'Live events daily • Free to register'. On the right side, there is a red button with the text 'Register now!'. The background of the banner is a collage of images including a person's face, a laptop, and abstract digital network patterns.

**ECS** The Electrochemical Society  
Advancing solid state & electrochemical science & technology

**239th ECS Meeting with IMCS18**

DIGITAL MEETING • May 30-June 3, 2021

Live events daily • Free to register

**Register now!**

# Detecting Spatial Autocorrelation for a Small Number of Areas: a practical example

Aswi Aswi<sup>1\*</sup>, Susanna Cramb<sup>2,4</sup>, Earl Duncan<sup>3</sup> and Kerrie Mengersen<sup>2,3</sup>

<sup>1</sup> Statistics Department, Universitas Negeri Makassar, Makassar, Indonesia

<sup>2</sup> Centre for Data Science, Queensland University of Technology, Brisbane, Australia

<sup>3</sup> School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

<sup>4</sup> Centre for Healthcare Transformation, School of Public Health and Social Work, Queensland University of Technology, Australia

**email:** \*aswi@unm.ac.id

**Abstract.** Moran's  $I$  is commonly used to detect spatial autocorrelation in spatial data. However, Moran's  $I$  may lead to underestimating spatial dependence when used for a small number of areas. This led to the development of Modified Moran's  $I$ , which is designed to work when there are few areas. In this paper, both methods will be presented. Many R programs enable calculating Moran's  $I$ , but to date, none have been available for calculating Modified Moran's  $I$ . This paper aims to present both methods and provide the R code for calculating Modified Moran's  $I$ , with an application to a case study of dengue fever across 14 regions in Makassar, Indonesia.

**Keywords:** Moran's  $I$ , Modified Moran's  $I$ , Spatial Autocorrelation, R Program

## 1. Introduction

Everything with a geographic location inevitably creates a spatial pattern [1]. Spatial analysis is used to analyse data related to geographic location. Spatial data can be in the form of discrete or continuous data and consists of three types, namely: areal data (also known as lattice data), geostatistical data (also known as point-reference), and point patterns [2]. Areal data are based on administrative boundaries such as regions, districts, counties, municipalities, postcode sections, and census tracts. Geostatistical data are suitable when data are observed at point locations. When focusing on the occurrence of an event and location are random, spatial point patterns are suitable.

Moran's  $I$  statistic proposed by Moran [3] in 1950 is one of the global indexes of spatial autocorrelation which is commonly used as a first step to detect spatial dependence in data [4]. Other global indexes are the Gamma index of spatial autocorrelation [5], Geary's  $C$  [6], Getis and Ord's  $G$  [7], and Join Count Statistics [8]. Moran's  $I$  measures the strength of spatial autocorrelation with values ranging from -1 to 1 [9]. Several modified versions of Moran's  $I$  have been suggested to improve its power [10-13]. However, none of these modifications address the case when there are a small number of areas. Moran's  $I$  may lead to underestimating spatial dependence when used for a small number of areas [14, 15]. This led to the development of a modified Moran's  $I$  (MMI) [14] which is designed to work when there are few areas.



Many R programs enable calculating Moran's  $I$ , but to date, none have been available for calculating MMI. This paper aims to present a practical example of both Moran's  $I$  and MMI methods, including a demonstration of the R code for calculating MMI, and a case study of dengue fever across 14 regions in Makassar, Indonesia.

## 2. Material

### 2.1. Moran's $I$

Moran's  $I$  can be used to measure spatial dependence among attribute values of geographic units (polygons or points). Without loss of generality, the following assumes areal units for illustration. The formula is defined as follows:

$$I = \frac{n}{\sum_{i=1}^n \sum_{k=1}^n w_{ik}} \frac{\sum_{i=1}^n \sum_{k=1}^n w_{ik} (y_i - \bar{y})(y_k - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1)$$

where  $y_i$  and  $y_k$  represent the attribute values at areas  $i$  and  $k$ ,  $n$  is the number of areas,  $\bar{y}$  is the average of the  $y$  values over the  $n$  areas, and  $w_{ik}$  is the  $i$ - $k$ <sup>th</sup> element of the spatial weights matrix,  $\mathbf{W}$ , which measures the spatial dependence between areas  $i$  and  $k$  [9, 16].

Moran's  $I$  formula can be written in notation matrix as follows:

$$I = \frac{n}{\text{sum}(\mathbf{W})} \frac{\mathbf{z}'\mathbf{W}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$$

where  $\mathbf{z} = \mathbf{y} - \bar{y}$ .

If the  $\mathbf{W}$  matrix is row standardised,  $\sum_{i=1}^n \sum_{k=1}^n w_{ik} = n$ , then Moran's  $I$  simplifies to:

$$I = \frac{\mathbf{z}'\mathbf{W}_r\mathbf{z}}{\mathbf{z}'\mathbf{z}}$$

### 2.2. Modified Moran's $I$ (MMI)

A modified Moran's  $I$  (MMI) that can detect spatial dependence even for very few areas was proposed by Carrijo and Da Silva [14] and is given as follows:

$$I_{\text{Mod}} = \frac{\sum_{i=1}^n (y_i - \bar{y}) (\sum_{k=1}^n w_{ik} y_k - \bar{y})}{[\sum_{i=1}^n (y_i - \bar{y})^2]^{1/2} [\sum_{i=1}^n (\sum_{k=1}^n w_{ik} y_k - \bar{y})^2]^{1/2}} \quad (2)$$

or in matrix notation:

$$I_{\text{Mod}} = \frac{\mathbf{z}'\mathbf{z}_w}{(\mathbf{z}'\mathbf{z})^{1/2} (\mathbf{z}_w'\mathbf{z}_w)^{1/2}}$$

where  $\mathbf{z} = \mathbf{y} - \bar{y}$ ;  $\mathbf{z}_w = \mathbf{W}_r^T \mathbf{y} - \bar{y}$ ; and  $\mathbf{W}_r^T$  is the row-standardised spatial weights matrix transposed. Based on equations (1) and (2), we can see that the difference between Moran's  $I$  and MMI is that in MMI formulae (equation 2), both the numerator and denominator consider neighbouring structure, while in Moran's  $I$  only the numerator considers the neighbouring structure. Another difference is that Moran's  $I$  is asymptotically normally distributed [17] while MMI is Student's  $t$  distributed [14].

### 2.3. Spatial neighbouring matrix

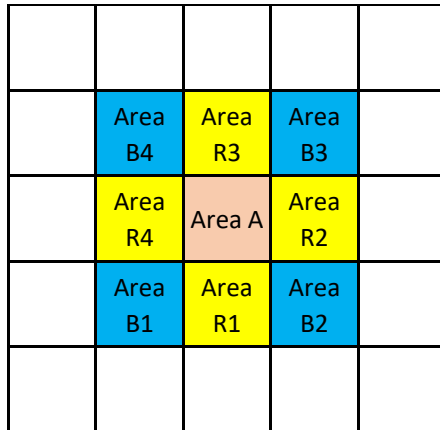
For areal data, the simplest spatial weights matrix is defined by the binary neighbourhood matrix as follows [16]:

$$w_{ik} = \begin{cases} 1 & \text{if areas } i \text{ and } k \text{ share a boundary} \\ 0 & \text{otherwise.} \end{cases}$$

For point referenced data, other criteria are used to determine whether points are neighbours. The resulting matrix is necessarily symmetric, that is  $w_{ik} = w_{ki}$ , and the constraint  $w_{ii} = 0$  is imposed.

The concept of a neighbourhood matrix is important in exploring areal data. Different forms of spatial neighbourhood matrices include rook contiguity (common border), bishop contiguity (common vertex),

and queen contiguity (common border and vertex) [9]. Explanation visual illustration of these spatial proximities is given as follows:



**Figure 1.** Rook, Bishop, and Queen Contiguity

Based on Figure 1, if area A is the area of interest, by using the first-order neighbourhood structure and rook contiguity, then the neighbouring areas are Area R1, R2, R3, and R4 (yellow). Using bishop contiguity, the neighbouring areas are Area B1, B2, B3, B4 (blue), while under Queen contiguity, all blue and yellow areas (R1 to R4 and B1 to B4) are considered neighbours. In this paper, we used the Queen contiguity neighbourhood structure, which is widely used.

### 3. Practical Example

#### 3.1. Moran's $I$

To show how Moran's  $I$  and modified Moran's  $I$  perform for a small number of areas, an example of a  $(3 \times 3)$  regular system with  $n = 9$  observations is given as follows:

155 #7	255 #8	155 #9
255 #4	405 #5	255 #6
155 #1	255 #2	155 #3

**Figure 2.** An example of random positive data using a regular system with  $n = 9$

Figure 2 shows an example of random positive data,  $y$ . The values in the lower-left corner depict the unique identifier for the enumeration spatial unit, and values in the centre depict the attribute values. By using the binary neighbourhood structure and Queen contiguity, the non-standardised spatial weights matrix is given as follows:

$$\mathbf{W} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$$

The row-standardised spatial weights matrix is given as follows:

$$\mathbf{W}_r = \begin{bmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 1/5 & 1/5 & 1/5 & 1/5 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/5 & 1/5 & 0 & 0 & 1/5 & 0 & 1/5 & 1/5 & 0 \\ 1/8 & 1/8 & 1/8 & 1/8 & 0 & 1/8 & 1/8 & 1/8 & 1/8 \\ 0 & 1/5 & 1/5 & 0 & 1/5 & 0 & 0 & 1/5 & 1/5 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/5 & 1/5 & 1/5 & 1/5 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0 \end{bmatrix}$$

where  $w_{r,ik} = \frac{w_{ik}}{\sum_k w_{ik}}$  and the elements of each row sum to one, and

$$\mathbf{z} = [\mathbf{y} - \bar{y}] = \begin{bmatrix} 155 - 227.22 \\ 255 - 227.22 \\ 155 - 227.22 \\ 255 - 227.22 \\ 405 - 227.22 \\ 255 - 227.22 \\ 155 - 227.22 \\ 255 - 227.22 \\ 155 - 227.22 \end{bmatrix} = \begin{bmatrix} -72.22 \\ 27.778 \\ -72.22 \\ 27.778 \\ 177.78 \\ 27.778 \\ -72.22 \\ 27.778 \\ -72.22 \end{bmatrix}$$

Moran's  $I$  can be calculated as follows:

$$I = \frac{[\mathbf{y} - \bar{y}]' \mathbf{W}_r [\mathbf{y} - \bar{y}]}{[\mathbf{y} - \bar{y}]' [\mathbf{y} - \bar{y}]} = \frac{-24444.44}{55555.56} = -0.44$$

The R library `ape` can be used to calculate Moran's  $I$  as follows (note that this package uses row-standardised spatial weights):

**Code 1.** R code for calculating Moran's  $I$  using the `ape` library

---

```
#Load required package
library(ape)

# Define the dummy data
y <- c(155, 255, 155, 255, 405, 255, 155, 255, 155)

# Define the spatial weights matrix
W <- matrix(c(
  0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0,
```

---

---

```

0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0,
1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1,
0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1,
0, 0, 0, 0, 1, 1, 0, 1, 0), 9, 9)

```

W

```
# Calculte Moran's I
```

```
I <- ape::Moran.I(y, W)
```

---

```
I
```

In practice, the spatial weights matrix is generally constructed from available spatial information, say from a shapefile. This can be achieved in R using the libraries `rgdal` and `spdep`.

The result of Moran's  $I$  as follows:

```

> I
$observed
[1] -0.44

$expected
[1] -0.125

$sd
[1] 0.1337339

$p.value
[1] 0.01850157

```

Alternatively, Moran's  $I$  can be calculated without using the `ape` library based on the matrix formulation as follows:

**Code 2.** R code for calculating Moran's  $I$  without using the `ape` library

---

```

# Calculate Moran's I first-hand
# First compute the row-standardised weights matrix
Wr <- W/(rowSums(W))

Z <- y - mean(y)
a <- t(Z) %*% Wr %*% Z
b <- t(Z) %*% Z
I.2 <- a/b
I.2

```

---

The result is as follows:

```

> I.2
      [,1]
[1,] -0.44

```

### 3.2. Modified Moran's $I$ ( $I_{Mod}$ )

By using the binary neighbourhood structure and queen contiguity, a column-standardised weight matrix ( $W_c$ ) and  $Z_w$  are given as follows:

$$\mathbf{z}_w = \mathbf{W}_r^T \mathbf{y} - \bar{y} = \begin{bmatrix} 0 & 1/5 & 0 & 1/5 & 1/8 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/5 & 1/8 & 1/5 & 0 & 0 & 0 \\ 0 & 1/5 & 0 & 0 & 1/8 & 1/5 & 0 & 0 & 0 \\ 1/3 & 1/5 & 0 & 0 & 1/8 & 0 & 1/3 & 1/5 & 0 \\ 1/3 & 1/5 & 1/3 & 1/5 & 0 & 1/5 & 1/3 & 1/5 & 1/3 \\ 0 & 1/5 & 1/3 & 0 & 1/8 & 0 & 0 & 1/5 & 1/3 \\ 0 & 0 & 0 & 1/5 & 1/8 & 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 1/5 & 1/8 & 1/5 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/8 & 1/5 & 0 & 1/5 & 0 \end{bmatrix} \begin{bmatrix} 155 \\ 255 \\ 155 \\ 255 \\ 405 \\ 255 \\ 155 \\ 255 \\ 155 \end{bmatrix} - \begin{bmatrix} 227.22 \\ 227.22 \\ 227.22 \\ 227.22 \\ 227.22 \\ 227.22 \\ 227.22 \\ 227.22 \\ 227.22 \end{bmatrix}$$

$$= \begin{bmatrix} -74.59722 \\ 28.73611 \\ -74.59722 \\ 28.73611 \\ 183.44444 \\ 28.73611 \\ -74.59722 \\ 28.73611 \\ -74.59722 \end{bmatrix}$$

MMI is therefore:

$$I_{\text{Mod}} = \frac{\mathbf{z}'\mathbf{z}_w}{(\mathbf{z}'\mathbf{z})^{1/2}(\mathbf{z}_w'\mathbf{z}_w)^{1/2}} = \frac{57355.56}{(235.70)(243.34)} = 0.9999998$$

The following code calculates the MMI in R.

### Code 3. R code for calculating MMI

```
# Compute modified Moran's I
Z<-(y-mean(y))
Zw <- t(Wr) %*% y - mean(y)
a <- t(Z) %*% Zw
b <- sqrt(t(Z) %*% Z)
c <- sqrt(t(Zw) %*% Zw)
I.mod <- (a/(b*c))
I.mod
```

Result is as follows:

```
> I.mod
[1,] 0.9999998
```

Using the specified weights matrix above, the value of Moran's  $I$  is negative (-0.44) while modified Moran's  $I$  (MMI) is positive ( $\approx 1$ ). The latter statistic provides a much better indication of the spatial autocorrelation of this data, whereas Moran's  $I$  indicates that the spatial autocorrelation is negative and not very strong in magnitude – an apparent contradiction to that which is readily observable [14].

### 3.3. Application on Dengue Fever dataset

The values of Moran's  $I$  and MMI using a dengue fever dataset in Makassar from 2002 and 2017 are provided below.

**Table 1.** The results of Moran's  $I$  and MMI for Makassar dengue fever data

Year	Moran's $I$	Modified Moran's $I$
2002	-0.3	-0.4
2003	-0.3	-0.4
2004	-0.2	-0.3
2005	-0.2	-0.3
2006	-0.2	-0.3
2007	-0.2	-0.3
2008	-0.2	-0.3
2009	-0.19	-0.24
2010	-0.2	-0.4
2011	-0.2	-0.2
2012	-0.1	-0.2
2013	-0.1	0.2
2014	-0.19	-0.25
2015	-0.1	-0.2
2016	0.14	0.19
2017	0.01	0.10

Table 1 shows that Moran's  $I$  is always smaller in magnitude than the MMI for each year. The largest difference between these two statistics is observed in 2013, where the value of Moran's  $I$  result is negative (-0.1) while the MMI value is positive (0.2). However, neither of these statistics suggests there is substantive spatial autocorrelation present in any year.

## 4. Conclusion

Moran's  $I$  can be misleading when there are a small number of areas. This can potentially affect the magnitude of the spatial autocorrelation and even the sign (negative/positive). This can indicate weak negative autocorrelation when the data has substantive positive clustering, for example. MMI is a valuable approach to obtain a more accurate measure of spatial autocorrelation when there are a small number of areas. The availability of the R code for calculating MMI is beneficial for researchers, especially for spatial modellers, and we hope this encourages greater usage of MMI.

## Acknowledgments

The authors acknowledge the City Health Department of Makassar for providing the dengue fever datasets.

## References

- [1] Klippel A, Hardisty F and Li R 2011 *Ann. Am. Assoc. Geogr.* **101** 1011-1031
- [2] Cressie N A C 1993 *Statistics for spatial data* (New York: Wiley)
- [3] Moran P A P 1950 *Biometrika* **37** 17
- [4] Lee J and Li S 2017 *Geogr. Anal.* **49** 36-57
- [5] Hubert L J, Golledge R G and Costanzo C M 2010 *Geogr Anal* **13** 224-233
- [6] Geary R C 1954 *The Incorporated statistician* **5** 115-146
- [7] Getis A and Ord J K 1992 *Geogr. Anal.* **24** 189-206
- [8] Mathur M 2015 *J. Appl. Nat. Sci.* **7** 501-513



- [9] Oyana T J and Margai F 2015 *Spatial analysis: statistics, visualization, and computational methods* (Boca Raton: CRC Press)
- [10] Jackson M C, Huang L, Xie Q and Tiwari R C 2010 *Int. J. Health Geogr.* **9** 33-33
- [11] Tango T 1998 *Stat. Med* **17** 1055-1058
- [12] Assunção R M and Reis E A 1999 *Stat. Med* **18** 2147-2162
- [13] Chen Y 2013 *PLoS ONE* **8** e68336-e68336
- [14] Carrijo T B and Da Silva A R 2017 *Geogr. Anal.* **49** 451-467
- [15] Aswi A, Cramb S, Duncan E and Mengersen K 2020 *Int J Health Geogr* **19** 1-39
- [16] Waller L A and Gotway C A 2004 *Applied spatial statistics for public health data* (Hoboken, N.J: John Wiley & Sons)
- [17] Tiefelsdorf M and Boots B 2016 *Environ. Plan A.* **27** 985-999