Geospatial Health 2025; volume 20:1277

Moran's *I* and Geary's *C*: investigation of the effects of spatial weight matrices for assessing the distribution of infectious diseases

Sarah Isnan,[1,2] Ahmad Fikri bin Abdullah,[3] Abdul Rashid Shariff,[1] Iskandar Ishak,[4] Sharifah Norkhadijah Syed Ismail,[5] Maheshwara Rao Appanan[6]

[1]Department of Biological and Agricultural Engineering, Faculty of Engineering, University Putra Malaysia, Serdang; [2]Department of Maritime Technology, Faculty of Defence Science and Technology, National Defence University of Malaysia, Cheras, Kuala Lumpur; [3]International Institute of Aquaculture and Aquatic Sciences (I-AQUAS), Universiti Putra Malaysia, Port Dickson; [4]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang; [5]Department Environmental and Occupational Health, Faculty of Medicine and Health Sciences, University Putra Malaysia, Serdang, Selangor; [6]Ministry of Health Malaysia, Putrajaya, Malaysia

Correspondence: Iskandar Ishak, Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Malaysia. E-mail: iskandar_i@upm.edu.my

Key words: Moran's *I*, Geary's *C*, spatial autocorrelation, COVID-19.

Contributions: SI, writing - review & editing, writing - original draft, visualization, validation, software, resources, project administration, methodology, investigation, funding acquisition, conceptualization; AfbA, conceptualization, supervision, critical revision of the manuscript for important intellectual content, final manuscript approval; ARS, conceptualization, supervision, critical revision of the manuscript for important intellectual content, final manuscript approval; II, conceptualization, supervision, critical revision of the manuscript for important intellectual content, final manuscript approval; SNSI, conceptualization, supervision, critical revision of the manuscript for important intellectual content, final manuscript approval; MRA, project administration, expert from Ministry of Health Malaysia, conceptualization, supervision, critical revision of the manuscript for important intellectual content, final manuscript approval.

Funding: this research was funded by Universiti Putra Malaysia, grant number 9738200.

Conflicts of interest: the authors declare no conflict of interest.

Ethical approval: the dataset used in our study is not publicly available due to privacy concerns. It contains sensitive information, specifically the latitude and longitude of the home addresses of COVID-19 positive patients, rounded to three decimal places. Ethical clearance is required to access this dataset. Our study was conducted under the ethical approval of the National Medical Research Register (NMRR) in Malaysia, with the NMRR ID: 22-01690-MBN and it has been approved on 3 June 2022. This approval confirms that our research adheres to the necessary ethical standards and guidelines. The data for our research was provided by the Ministry of Health Malaysia, following a formal request from our team. The data collection process involved anonymized records, which did not necessitate individual informed consent as per the ethical guidelines outlined by the NMRR.

Availability of data and material: the data that support the study's findings are available from Ministry of Health Malaysia, but there are restrictions on their availability because they were used under license for the current study and thus are not publicly available. However, the authors' data are available upon reasonable request and with the permission of Ministry of Health Malaysia.

Acknowledgement: the authors would like to acknowledge the support from University Putra Malaysia under IPS-GRANT (ID Number: 9738200).

Received: 26 February 2024.
Accepted: 16 September 2024.

Geospatial Health 2025; 20:1277
doi:10.4081/gh.2025.1277

Abstract

The COVID-19 outbreak has precipitated severe occurrences on a global scale. Hence, spatial analysis is crucial in determining the relationships and patterns of geospatial data. Moran's *I* and Geary's *C* are prominent methodologies used to measure the spatial autocorrelation of geographical data. Both measure the degree of similarity or dissimilarity between nearby locations based on attribute values in such a way that the selection of distance techniques and weight matrices significantly impact the spatial autocorrelation results. This paper aimed at carrying out the spatial epidemiological characteristics analysis of the pandemic comparing the results of Moran's *I* and Geary's *C* with different parameters to gain a comprehensive understanding of the spatial relationship of COVID-19 cases. We employed distance-based techniques, K-nearest neighbour, and Queen contiguity techniques to assess the sensitivity of the different parameter configurations for both Moran's *I* and Geary's *C*. The findings revealed that former provided more reliable and robust results compared to the latter, with consistent results of spatial autocorrelation (positive spatial autocorrelation). The distance weight of 0.05 using the Manhattan method of Moran's *I* is the recommended distance weight, as it outperformed other weight matrices (Moran's I = 0.0152, Z-value=110.8844 and p-value=0.001).

Introduction

Spatial autocorrelation techniques such as Moran's Index (Moran's *I*) and Geary's Contiguity ratio (Geary's *C*) are vital for the determination of spatial patterns in a dataset, such as the distribution and pattern of disease outbreaks, pandemics in particular. Both of these methods measure the similarity and dissimilarity of the neighbouring locations based on attribute values. With regard to infectious diseases, Moran's *I* has consistently shown significant positive spatial autocorrelation with different distance weight measurements indicating that areas with a high number of cases of diseases tend to be surrounded by other areas with similar high numbers. This is aligned with the characteristics of pandemics, for example the recent pandemic of the coronavirus disease 2019 (COVID-19), where nearby areas have been more likely to have similar infection rates due to factors, such as the patterns of social interaction, various environmental factors and population density (Silalahi *et al.*, 2020). Kuala Lumpur, Gombak and Petaling represent Malysian areas with a high total population (around

5,235,400 people) making these areas crowded and congested (Malaysia, 2022), something which increases the susceptibility to infection due to limited social distancing, high levels of physical contact, crowded living conditions and the strong probability of large outdoor gatherings of people (Ayouni *et al.*, 2021).

Moran's *I* is commonly applied for the study of widely spread diseases, e.g., COVID-19 and dengue, and papers reporting this have been published in a large number of countries, e.g., Kang *et al.*, (2020) in China, Wetchayont & Waiyasusri (2021) in Thailand, da Silva *et al.* (2022) in Brazil and Vilinová & Petrikovičová (2023) in Slovakia, with global implications reported by Morais & Gomes (2021). Suryowati *et al.* (2018) are among the few researchers using Geary's *C*, which they applied for the determination of the spatial autocorrelation of dengue cases in Indonesia. All of the studies cited here concerned local outbreaks with a strong potential for spread and default weight matrices were used in order to determine the spatial autocorrelation. Default settings are generally applied in the study of spatial autocorrelation but Chen (2021) emphasizes the limitation of this approach as it may not accurately capture the spatial association. Consistent, reliable and robust responses are highly important, particularly in the study of the spread of threatening pandemics, such as COVID-19. As it is crucial to enhance validity and reliability of the methodology used, we compared the sensitivity of Moran's *I* and Geary's *C* when applied with different weight matrices with the aim of identifying the most suitable weight matrix to determine spatial autocorrelation.

Materials and Methods

Dataset

In this study, we utilized a very detailed dataset consisting of daily COVID-19 positive cases recorded for almost three years and residential addresses expressed by latitude and longitude (data rounded to 3 decimal places. The high granularity of the dataset is crucial to allow for precise geospatial analysis and accurate mapping of COVID-19 transmission patterns The timeline of the dataset was from 4 February 2020 until 19 December 2022 which comprised a total of 1,0008,518 patients of positive COVID-19 cases. This big pool of data was felt to be necessary for accurate analysis within the study area that encompassed three districts situated in the western region of Peninsular Malaysia: Petaling, Gombak and Kuala Lumpur. These areas were chosen because they have the highest populations in Malaysia, including Petaling with 2,290,000 people, Kuala Lumpur with 2,000,000 people, and Gombak with 945,400 people (https://www.dosm.gov.my). Petaling presents unique challenges in controlling the spread of COVID-19 as it has the highest population density among these districts. Kuala Lumpur, the nation's capital, is also densely populated and serves as a major hub for nation's economic activities hence making it a vulnerable region for transmission of infectious diseases. Gombak, with a smaller population, but due to its close proximity to Kuala Lumpur, cross districts transmission patterns might occur in this district potentially propelling the number of COVID-19 cases to become high also in this region. Each COVID-19 cases are reported together with additional data, such as brand of vaccine taken, age, gender, symptomatic status and nationality. This study randomly selected a total of 1,00,8158 cases from the year of 2020 until 2022 to avoid biases that may arise (Shaham *et al.*, 2020). The year 2020 was chosen as the earliest outbreak due to COVID-19 appeared in Malaysia during that time. The data were recorded via MySejahetra. a COVID-19 monitoring mobile application used by the Ministry of Health, Malaysia. Figure 1 shows the study area map and Figure 2 depicts the total populations for each district. Figure 3 shows the distribution of the COVID-19 cases in Gombak, Petaling and Kuala Lumpur based on the location of residential addresses with the geographical coordinates rounded to 3 decimal places. Due to the large number of data points, which exceeds one million cases, the map shows a very densely populated view. The small scale of the map, combined with the sheer volume of data points, results in an overlapping and cluttered appearance, which might be perceived as visually overwhelming.

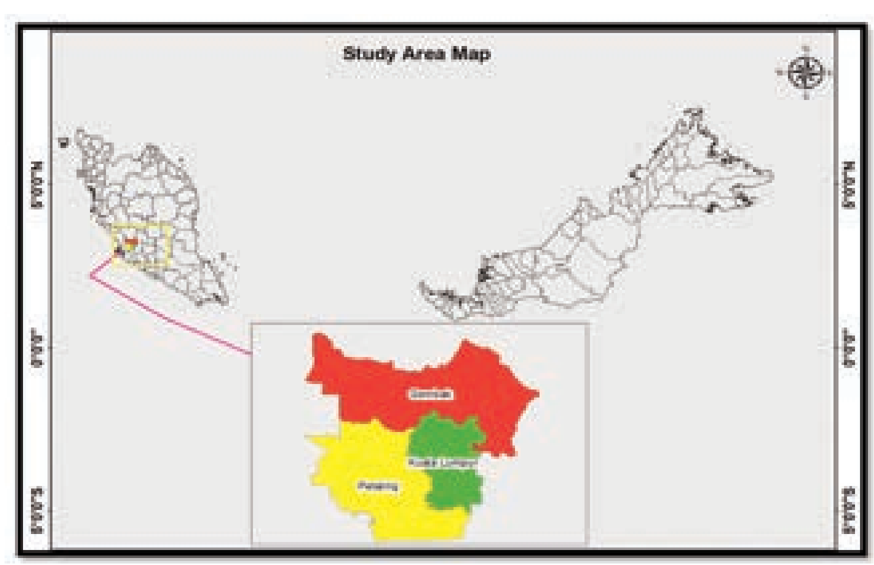

Figure 1. Map of the study area in Malaysia.

Statistics

Getis-Ord-Gi* (1992) is one of the spatial statistics techniques used to identify spatial clusters and hotspots and coldspots in the dataset. It is calculated as:

$$G_i^* = \frac{\sum_{j=1}^{n} \omega_{i,j} x_j - \bar{X}\sum_{j=1}^{n} \omega_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} \omega_{i,j}^2 \left(\sum_{j=1}^{n} \omega_{i,j}\right)^2\right]}{n-1}}} \qquad \text{Eq. 1}$$

where x_j is the attribute value for feature j; $w_{i,j}$ the spatial weight between features i and j; and n is equal to the total number of features:

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \qquad \text{Eq. 2}$$

where

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

The Gi* statistic is a *Z*-score, hence no further calculation is required. Moran's *I* (1950), on the other hand, is a measure of spatial autocorrelation expressed as,

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w^{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{Eq. 3}$$

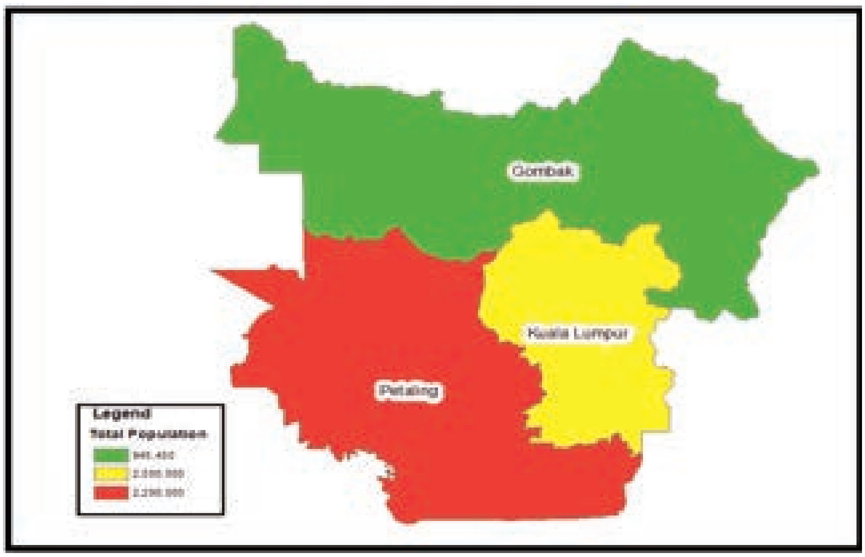

Figure 2. Population distribution at the district level.

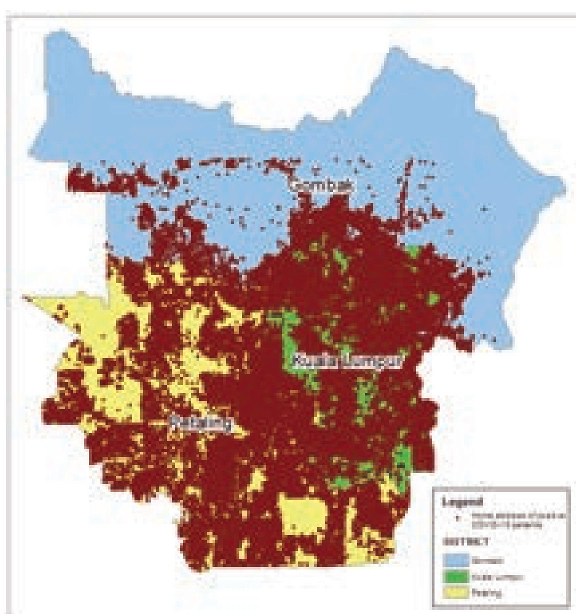

Figure 3. Distribution of the COVID-19 cases in Gombak, Petaling and Kuala Lumpur based on the location of residential addresses.

where n is the number of spatial units; x_i and x_j the values of the variables in spatial units; $\bar{x}$ the mean of the variable in all spatial units; and $w_{i,j}$ is the spatial weight between spatial units i and j.

A Moran's) close to 1 indicates a strong positive spatial autocorrelation, which is a sign that similar value tends to cluster together in the space investigated. Conversely, a Moran's) value closer to -1 indicates a strong negative spatial autocorrelation, which is a sign that dissimilar value tends to cluster in that space. Values around 0 indicate absence of a significant spatial autocorrelation, which is a sign of random spatial distribution values. In this study, we used the frequency of the numbers of positive COVID-19 cases occurring at each spatial unit as attribute data. To evaluate the statistical significance of Moran's *I*, we calculated the *Z*-score to assess its significance. The presence of a *Z*-score inside the critical region rejects the null hypothesis (H_0) suggesting that there is spatial autocorrelation in the dataset, while a *Z*-score outside the critical region indicates the absence of spatial autocorrelation in the dataset. The *Z*-score is calculated as:

$$Z = \frac{I - E\,[I]}{\sqrt{Var\,[I]}} \qquad \text{Eq. 4}$$

where I is the value of Moran's I; $E\,[I]$ the expected value of Moran's I for the null hypothesis; and $Var\,[I]$ the variance of Moran's I.

The *p*-value is calculated to test the hypothesis and assess the strength of the null hypothesis. It indicates that level of the chance that the dataset is statistically significant rather than random chance occurrence and is calculated as:

$$p = P\,(Z \geq |Z_{Observed}\,|) \qquad \text{Eq. 5}$$

where $Z_{Observed}$ is the Z-score of Moran's I; and $P\,(Z \geq |\,Z_{Observed}\,|)$ the probability of observing a *Z*-score being equal or higher than the observed one. A $p<0.05$ outcome indicates that the value of Moran's *I* is different from that expected under the null hypothesis resulting in the rejection of the null hypothesis as it shows the presence of spatial autocorrelation in the dataset, and vice versa.

Geary's C, finally, measures the spatial autocorrelation in a dataset through the formula:

$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{2W\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{Eq. 6}$$

where is the number of spatial units; and x_j the values of the variable in spatial units; the mean of the variables in all spatial units; w_{ij} is the spatial weight between the spatial units i and j; and W the sum of the spatial weights (w_{ij}).

A value of Geary's *C* lower than 1 indicates positive spatial autocorrelation, a value higher than 1 negative spatial autocorrelation and values close to 1 absence of significant spatial autocorrelation. We applied the Moran's *I* and Geary's *C* to the datasets using various parameter settings as shown in Table 1. In addition, 999 permutations for Euclidian, Manhattan and Queen contiguities were applied to enhance the robustness of the results. All programming was run in JupyterNotebook, a web-based (https://jupyter.org/) computing platform. The summary of the processing steps to determine Moran's *I*, Geary's *C*, *p*-values and Z-scores is shown in Figure 4.

Proximity metrics

The Euclidian distance is a measure of two points in a straight line. For example, the position of two points are (x_1, y_1) and (x_2, y_2), hence the Euclidian distance of two points can be calculated (Dokmanic *et al.*, 2015) as:

$$Euclidian\ Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The Manhattan distance is a measure of the distance of two points in a grid-based system. For example, the position of two points are (x_1, y_1) and (x_2, y_2), hence the Manhattan distance of two points can be calculated (Malkauthekar, 2013) as:

$$Manhattan\ Distance = |x_2 - x_1| + |y_2 - y_1|$$

K-nearest neighbour is a supervised machine-learning method used for classification and regression (Rahman *et al.*, 2021). The default method to assign weight for k-nearest neighbour is Euclidian distance. In order to determine the optimum number of neighbours (k), the Elbow curve method (Thorndike, 1953) was performed and the results show that k=5 is the maximum number of neighbours. The spatial weight for given data points is expressed as:

$$\begin{cases} 1, if\,i \neq j\,and\,d_{ij} \leq d_k \\ 0, otherwise \end{cases} \qquad \text{Eq. 7}$$

where w_{ij} is the spatial weight between spatial units i and j; $d_{i,j}$ the Euclidian distance between i and j; d_k the threshold distance, which determines how many points are neighbours. If $d_{i,j}$ £ d_k, both points are considered neighbours; otherwise not. Hence, the formula above assigns the weight = 1 to pairs of data within the threshold distance, which indicates neighbour relationship, while pairs that are not neighbours are assigned the weight = zero.

The Euclidian distance technique measures the distance in a straight line between two points, while the Manhattan one measures the sum of absolute differences in coordinates in both vertical and horizontal axes. Due to this, the latter technique tends to be more sensitive to local spatial patterns, especially when using

Table 1. Weight matrices of Moran's *I* and Geary's *C*.

Method	Distance technique	Distance type
Inverse Distance	Euclidean	0.01 0.02 0.03 0.04 0.05
Inverse distance squared	Manhattan	0.01 0.02 0.03 0.04 0.05
K-nearest neighbour	Nil	k=8 k=9 k=10 k =15
Queen contiguity	Nil	Nil

smaller distance weights, which results in higher Moran's *I* values (Syetiawan *et al.*, 2022). The Queen contiguity matrix method considers two spatial units as neighbours when they share a boundary or vertex. The calculation of the method for a binary spatial weights' matrix is express as:

$$\begin{cases} 1, if spatial units i and j are neighbours \\ 0, otherwise \end{cases}$$

where w_{ij} is the spatial weight between spatial units *i* and *j*. If *i* and *j* share common boundaries or vertex, then $w_{ij} = 1$; otherwise = 0.

Results

Supplementary Materials, Table 1 shows the result of the spatial autocorrelation test. All four methods using Moran's *I* with different weight matrices consistently produced significant positive spatial autocorrelation results. However, Geary's C yielded inconsistent results with different weight matrices. Significant negative spatial autocorrelation was observed using a 0.01 distance weight with Manhattan, Euclidian and K-nearest neighbour, while other distance weights produced insignificant positive spatial autocorrelation results.

The distance weight of 0.01, *i.e.*, analysis at a very local scale, suggests that neighbouring locations are less similar to each other, which indicates a sign of dispersion (Juliani & Nasution, 2024). It shows that the weighted sum of squared differences among the adjacent locations was high compared to the global variance. Conversely, the distance weight of ≥0.02 produced insignificant positive spatial autocorrelation suggesting that nearby locations were more similar to each other, with the weighted sum of squared differences between adjacent location low compared to the global variance (Saiful Bahri *et al.*, 2014).

Figure 5 illustrates the result of Moran's *I* with the Euclidian and Manhattan techniques. As expected, the latter method yielded higher result with Moran's *I* results compared to the former. However, as seen in Figure 6, the results were the opposite when Geary's *C*. and as the distance weight increases, the Geary's *C* values for both techniques also increases indicating more similarity between the neighbours as noted by Chen (2021). Figures 7 and 8 show that Moran's *I* results decrease when the distance weight increases. Conversely, when the distance weight increases, Geary' *C* results also increase showing that this method is more sensitive

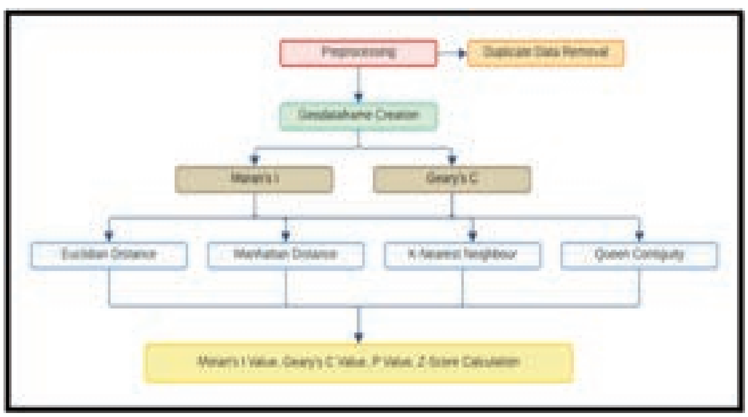

Figure 4. Process summary.

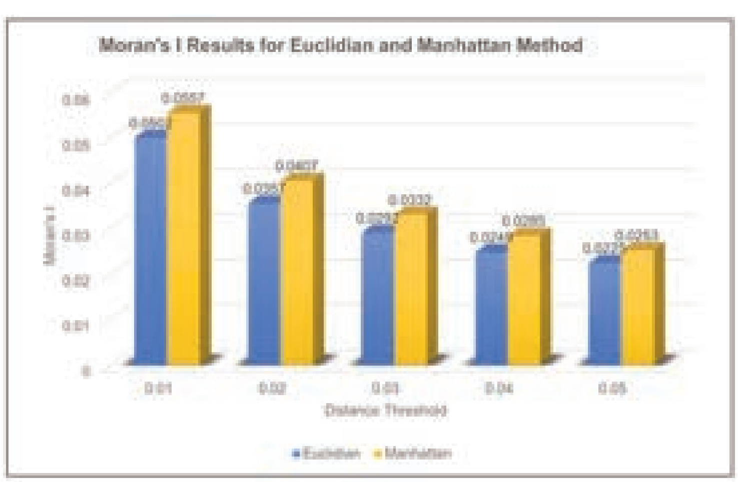

Figure 5. Results of Moran's *I* based on the Euclidian and Manhattan methods.

when larger distance weights are involved (Chen, 2021).

Figure 9 illustrates the results of the k-nearest neighbour method *C* using different numbers of neighbours (k). The results indicate that when the number of neighbours increases, the results of Moran's *I* and Geary's *C* both decrease. As seen in the Figure Moran's *I* values are lower than those of Geary's *C*, i.e. the latter is more sensitive to local spatial patterns when the Queen contiguity method is used (Chen, 2021).

An overview over the sensitivity, and potential applications in spatial autocorrelation analysis of Moran's *I* of Geary's *C* with special reference to advantages under various circumstances is shown in Table 2.

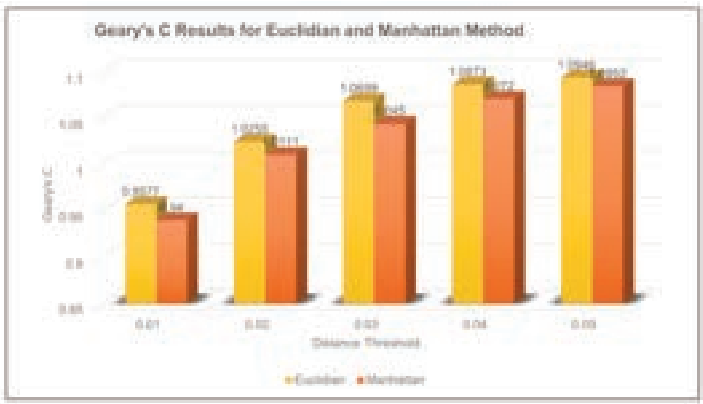

Figure 6. Results of Geary's *C* based on Euclidian and Manhattan methods.

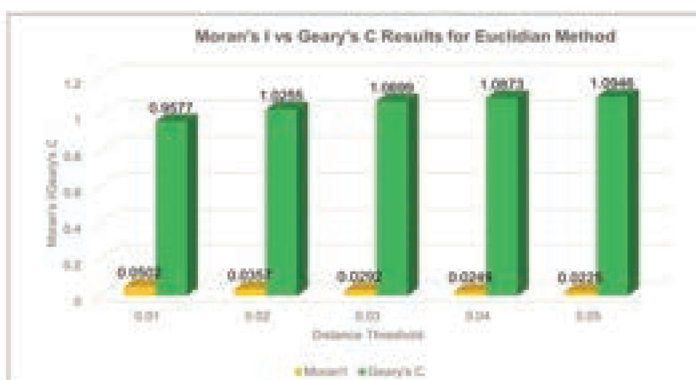

Figure 7. Results of Moran's *I* *vs* Geary's *C* with the Euclidian method.

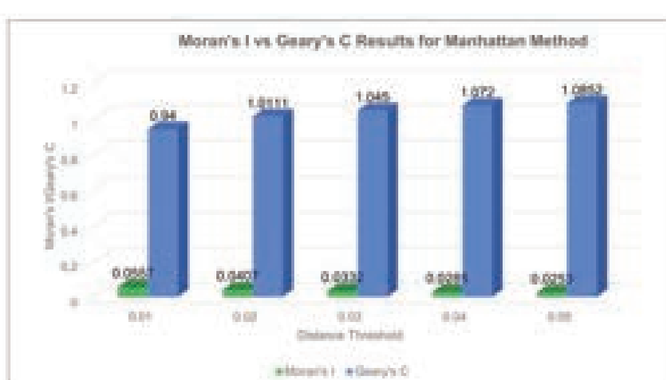

Figure 8. Results of Moran's *I* *vs* Geary's *C* with the Euclidian method.

Discussion

As the distance weight increases, the Moran's *I* values decreases which is aligned with Tobler's (1970) first law of geography, which states that "Everything is related to everything else, but near things are more related than distant things". This pattern is evident across both Euclidean and Manhattan distance techniques, the results of which reflect diminishing spatial autocorrelation when the analysis considers higher spatial scales (Saiful Bahri *et al.*, 2014). However, the Manhattan distance technique produced slightly higher Moran's *I* values compared to the Euclidian technique at similar distance weights (Malkauthekar, 2013), a fact that can be attributed to the nature of the Manhattan distance, which sums the absolute differences in coordinates along both vertical and horizontal directions (Li *et al.*, 2007). The increased sensitivity for local spatial patterns is practically useful in determining clustering in urban environments, where the spatial layout follows a grid pattern, as opposed to the straight-line Euclidian distance (Ei *et al.*, 2023).

Geary's *C*, on the other hand, produces a different perspective on the spatial autocorrelation by focusing more on local variations rather than global patterns, which highlights its sensitivity to the choice of spatial weight matrices and distance techniques. This is due to a particular responsiveness to local changes and Geary's *C* is therefore more likely to detect areas of spatial heterogeneity (Chen, 2021) making it valuable for detecting localized spatial anomalies or areas where the spatial relationship between points differs from overall trends (Chen, 2023). Nevertheless, this sensitivity also means that Gearys' *C* produces less stable and inconsistent results when applied across different techniques and different distance weights.

The k-nearest neighbour methods offers flexibility by defining spatial relationships based on the number of nearest neighbours rather than fixed distances (Bangira *et al.*, 2019). We found that increasing the number of neighbours reduced the spatial autocorrelation value for both Geary's *C* and Moran's *I*, something that moved the results towards spatial dispersion. Thus, the spatial autocorrelation effect drops as more distant and unrelated points are considered during the analysis (Rahman *et al.*, 2021). Hence, a very high precaution shall be taken when using this technique, especially in epidemiology dealing with high rates of bacterial and viral transmissions. Additionally, the Queen contiguity method, which defines neighbours based on shared boundaries, also showed significant positive spatial autocorrelation with both techniques. This suggests strong clustering outcomes, which was especially evident in the densely populated areas in Gombak, Petaling and Kuala Lumpur.

The comparison of Moran's *I* and Geary's *C* with different

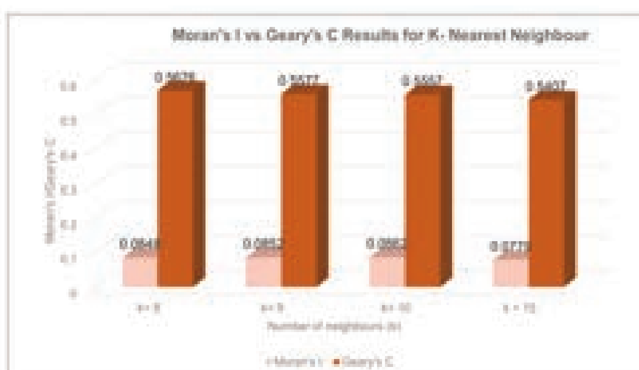

Figure 8. Results of Moran's *I vs* Geary's *C* with the k-nearest neighbour method.

Table 2. Overview the characteristics of Moran's *I* and Geary's *C*.

Characteristic	Distance technique	Distance type
Types of measure	Global spatial autocorrelation	Global spatial autocorrelation
Function	Measures global covariance (clustering of similar values)	Measures global differences (dissimilarity between neighbours)
Mathematical range	-1 to +1	0 to 2
Interpretation	Positive values indicate clustering while negative values indicate dispersion	Values less than 1 indicate positive autocorrelation (similarity) while valuesmore than 1 indicates negative autocorrelation (dissimilarity)
Sensitivity	Less sensitive to local variations and more suitable for global patterns	More sensitive to local variations and measures local differences
Advantages	Provides a broad overview of spatial patterns	Sensitive to local changes
Disadvantages	Might overlook local variations	Results might be inconsistent with varying parameters
Sensitivity to parameters	Generally robust across different spatial weight matrices	Highly sensitive to changes in spatial weight matrices and other parameters
Application examples	Clustering of disease cases, socio-economic data	Detection of spatial outliers, localized patterns of inequality
Preferable approach	Manhattan distance for detecting clustering	Euclidean distance for detecting local dissimilarities

parameters provides a profound understanding of spatial autocorrelation as it leads to the identification of parameter settings producing consistent results. It also highlights the potential biases and limitations correlated with specific parameter choices (Phillips *et al.*, 2020). In our study, the most suitable method for the dataset used was found to be the inverse distance approach based on the Manhattan distance using a weight of 0.05 as this distance weight produced a positive Moran's *I* value with a high *Z*-value and low *p*-value (Supplementary Materials, Table 1). The consistency across different distance weights and techniques makes it a robust choice to understand the global spatial patterns, while Geary's *C* offers a more nuanced approach by emphasizing local spatial variations. The different sensitivity of the Geary's *C* with respect to different weight matrices would make it useful in potential studies in identifying local anomalies providing support for targeted interventions, such as local diagnostic programmes and vaccination approaches (Ei *et al.*, 2023).

The k-nearest neighbour method can be utilised in regions where population density varies widely as it ensures that a consistent number of neighbours is analysed regardless of the spatial context. Nonetheless, this method should be utilised with precaution, especially in local areas with high transmission as higher number of neighbours may produce lower values for spatial autocorrelation and thus produce false results. The Queen contiguity is suitable to be utilised in urban areas where spatial clustering is closely intertwined. Together with physical proximity, this technique highlights the spatial boundaries and contiguity consideration in epidemiological studies.

Limitations

One of the limitations of this study is that the dataset consisted of reported cases from the Malaysian Ministry of Health, which means that the presence of unreported cases would create a bias, the strength of which remains unknown. This study also considered spatial autocorrelation based on the frequency of the COVID-19 cases with respect to location (the spatial element), while the additional consideration of factors, such as those based on socio-economy, detailed demography and health interventions could have changed outcomes. Future studies should also focus on other factors contributing to transmission, e.g., vaccination based on the fact that different vaccines used against COVID-19 have varying levels of protection. By studying the spatial autocorrelation of vaccination types together with the distribution of COVID-19 cases, it would be possible to understand the impact of vaccination interventions in mitigating and controlling transmission. It is crucial to provide all information of potential impact. such as additional data would provide a better understanding of the factors driving spatial autocorrelation.

Conclusions

We compared the result of spatial autocorrelation with different parameter settings to allow a deeper insight on the spatial relationships by taking account the sensitivity of the weight matrices. By using a high-resolution dataset, this study should provide more accurate and reliable spatial patterns and dynamics of disease outbreaks. Compared to other weight matrices, a Manhattan distance weight of 0.05 with Moran's *I* was found to be the most suitable for the dataset used.

References

Ayouni I, Maatoug J, Dhouib W, Zammit N, Fredj SB, Ghammam R, Ghannem H, 2021. Effective public health measures to mitigate the spread of COVID-19: a systematic review. BMC Public Health 21:1015.

Bangira T, Alfieri SM, Menenti M, van Niekerk A, 2019. Comparing thresholding with machine learning classifiers for mapping complex water. Remote Sens 11:1351.

Chen Y, 2021. An analytical process of spatial autocorrelation functions based on Moran's index. PLoS ONE 16:0249589

Chen Y, 2023. Spatial autocorrelation equation based on Moran's index. Sci Rep 13:1–14.

da Silva CFA, Silva MC, dos Santos AM, Rudke AP, do Bonfim CV, Portis GT, de Almeida Jr PM, Coutinho MBS, 2022. Spatial analysis of socio-economic factors and their relationship with the cases of COVID-19 in Pernambuco, Brazil. Trop Med Int Health 27:397–407.

Dokmanic I, Parhizkar R, Ranieri J, Vetterli M, 2015. Euclidean Distance Matrices: Essential theory, algorithms, and applications. IEEE Signal Process Mag 32:12–30.

Ei SU, Laohasiriwong W, Sornlorm K, 2023. Spatial autocorrelation and heterogenicity of demographic and healthcare factors in the five waves of COVID-19 epidemic in Thailand. Geospat Health 18:1183

Getis A, Ord JK, 1992. The Analysis of Spatial Association by Use of Distance Statistics. Geogr Anal 24:189–206.

Juliani IP, Nasution H, 2024. Spatial autocorrelation analysis using the moran and lisa index on the spread of malaria disease in North Sumatra Province. Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika 5;154–64.

Kang D, Choi H, Kim JH, Choi J, 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. International J Infect Dis 94:96–102.

Li H, Calder CA, Cressie N, 2007. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. Geogr Anal 39:357–75.

Malkauthekar MD, 2013. Analysis of Euclidean distance and Manhattan distance measure in face recognition. IET Conf Publ CP646:503–07.

Morais LR de A, Gomes GS da S, 2021. Applying spatio-temporal scan statistics and spatial autocorrelation statistics to identify COVID-19 clusters in the world-a vaccination strategy? Spatial Spatio-Temp Epidemiol 39:100461.

Moran P, 1950. Continuous Stochastic Phenomena. Biometrika 37:17–23.

Phillips B, Anand M, Bauch CT, 2020. Spatial early warning signals of social and epidemiological tipping points in a coupled behaviour-disease network. Sci Rep 10:1–12.

Rahman MS, Pientong C, Zafar S, Ekalaksananan T, Paul RE, Haque U, Rocklöv J, Overgaard HJ, 2021. Mapping the spatial distribution of the dengue vector Aedes aegypti and predicting its abundance in northeastern Thailand using machine-learning approach. One Health 13:100358

Saiful Bahri MA, Karbassi Z, Majid MR, Sabri S, Johar F, Ludin, AMN. 2014. Comparison of Spatial Autocorrelation Analysis Methods for Distribution Pattern of Diabetes Type II Patients in Iskandar Malaysia Neighbourhoods. In Int Conf Urban Reg Planning: Planning in Uncertain World, 9-11 May, 2014, Skudai, Johor. Malaysia. pp. 1–18.

Silalahi FES, Hidayat F, Dewi RS, Purwono N, Oktaviani N, 2020.

GIS-based approaches on the accessibility of referral hospital using network analysis and the spatial distribution model of the spreading case of COVID-19 in Jakarta, Indonesia. BMC Health Serv Res 20:1053.
Suryowati K, Bekti RD, Faradila A, 2018. A comparison of weights matrices on computation of dengue spatial autocorrelation. IOP Conference Series: Materials Sci Engineer 335:012052
Syetiawan A, Harimurti M, Prihanto Y, 2022. A spatiotemporal analysis of COVID-19 transmission in Jakarta, Indonesia for pandemic decision support. Geospat Health 17:85–97.
Thorndike RL, 1953. Who belongs in the family? Psychometrika 18:267-76
Vilinová K, Petrikovičová L, 2023. Spatial autocorrelation of COVID-19 in Slovakia. Trop Med Infect Dis 8:298.
Wetchayont P, Waiyasusri K, 2021. Using Moran'S I for Detection and Monitoring of the Covid-19 Spreading Stage in Thailand During the Third Wave of the Pandemic. Geogr Environ Sustain 14:155–67.

Online supplementary materials

Table 1. Moran's I and Geary's C with different weight matrices.