World Journal of Biology Pharmacy and Health Sciences
eISSN: 2582-5542
Cross Ref DOI: 10.30574/wjbphs
Journal homepage: https://wjbphs.com/

(RESEARCH ARTICLE)

Check for updates

Use of Getis-Ord's statistic to detect hotspots and coldspots of COVID-19 in Hanoi City, Vietnam

Thi-Bich-Thuy Luong * and Thi-Hien Cao

Faculty of Nursing, East Asia University of Technology, Hanoi, Vietnam.

World Journal of Biology Pharmacy and Health Sciences, 2023, 15(03), 102–109

Publication history: Received on 08 August 2023; revised on 16 September 2023; accepted on 19 September 2023

Article DOI: https://doi.org/10.30574/wjbphs.2023.15.3.0394

Abstract

Background: The emergence and rapid spread of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), a potentially fatal disease, is swiftly leading to public health crises worldwide. This study aimed to detect hotspots and coldspots of COVID-19 in Hanoi city, Vietnam.

Methods: the Getis-Ord's G_i^* statistic-based hotspot analysis was employed to detect hotspots and coldspots of COVID-19 pandemic in Hanoi city. Two methods for constructing spatial weight matrix have been used, namely the first and second order of contiguity.

Results: it was found from a case study of COVID-19 cases reported on 31 January 2022 in Hanoi city, a total of six hotspots and six coldspots of COVID-19 were detected using the first order contiguity (statistically significance at the 0.05 level). For the case of using the second order of contiguity, six hotspots and three coldspots were successfully identified. Hotspots were mainly concentrated in urbarn districts in the east of Hanoi. Coldspots were detected in the western suburban districts.

Conclusions: the study results has proven the effective use of Getis-Ord's statistics to detect hotspots and coldspots of COVID-19 pandemic. Findings in this study make great contributions to our understanding of the spatial clustering of the COVID-19 pandemic.

Keywords: Hotspots; Coldspots; Getis-Ord's G_i^* statistic; COVID-19; Hanoi; Vietnam

1. Introduction

Since early 2020, the COVID-19 pandemic has posed a major risk to public health on a global scale. The COVID-19 pandemic was brought on by SARS-CoV-2, the coronavirus that causes severe acute respiratory syndrome (1). Therefore, Eastern Asia, Europe, and the rest of the world were quickly infected by the virus (2). According to the most recent data, the World Health Organization had reports of more than 770.4 million confirmed cases of COVID-19 as of September 13, 2023, including more than 6.9 million fatalities (3). With the increasing availability of high-quality data and improved computational capabilities, numerous geospatial methods and tools have been developed and used in infectious diseases, including COVID-19 surveillance (4,5). As a result, attempts have been made to examine the COVID-19 pandemic, especially in studies of the COVID-19 epidemic's hotspots (6,7).

In spatial epidemiology, spatial clustering analysis plays an important role in identifying spatial aggregation of disease cases by identifying whether geographically grouped cases can be explained by chance or are statistically significant to find evidence of etiologic factors (8,9). Previous research have shown that several social processes connected to the

* Corresponding author: Thi-Bich-Thuy Luong; Email: thuylb@eaut.edu.vn

space of occurrence frequently determine the dispersion of infectious diseases (10,11). COVID-19 infections may vary in both space and time due to the complex interactions of various influences, including socioeconomic vulnerability, rapid population increase, and urbanization, as well as environmental factors (12). To conduct a study of outbreaks, spatial analysis and the identification of areas with COVID-19 clusters, followed by the characterisation of the causes of the dynamics in these clusters, have been encouraged (13,14). The resulting maps from these spatial methods can help prevent and control cases with targeted public health action plans and guided interventions in areas with higher than expected disease risk while motivating the population with various public health programs with the advanced knowledge of disease etiological characteristics (14). The most frequently used method for spatial clustering and hotspot analysis was local Moran's I, followed by *Getis-Ord's* G_i^* statistic, Kulldorff's spatial scan statistic and Kernel density (15). For example, using *Getis-Ord's* G_i^* statistic and geographically weighted principal component analysis, the impact of living environment deprivation on COVID-19 hotspot was successfully examined in Kolkata megacity, India (16). the exploratory spatial data analysis and the geodetector method was employed to analyze the spatial and temporal differentiation characteristics and the influencing factors of the COVID-19 epidemic spread in mainland China based on the cumulative confirmed cases, average temperature, and socio-economic data (17). The global (Moran's I) and local indicators of spatial autocorrelation (LISA), both univariate and bivariate, were successfully applied to derive significant clustering of COVID-19 pandemic (18). The global Moran's I statistic and the retrospective space-time scan statistic were also successfully used to analyze spatio-temporal clusters of COVID-19 (19).

The objective of this study was to detect hotspots and coldspots of COVID-19 in Hanoi city, Vietnam. The *Getis-Ord's* G_i^* statistic-based hotspot analysis was used to identify hotspots and coldspots of COVID-19 pandemic in Hanoi city. Two methods for constructing spatial weight matrix have been used including the first order and second order of contiguity.

2. Material and methods

In this study, a dataset of COVID-19 cases in the 2022 Winter Day in Hanoi was used to analysis of spatial clusters. These COVID-19 locally transmitted cases were reported on 31 January 2022. The spatial distribution of these COVID-19 cases is shown in Figure 1. Data from Figure 1 illustrate that the COVID-19 cases were mainly reported in the northeast of Hanoi city. In particular, high numbers of COVID-19 case were mainly confirmed in the Hanoi metropolitan where the population density is dense. Data from Table 1 shows that high numbers of COVID-19 cases were reported mostly in urban districts in the northeast of Hanoi city. On the other hand, low and very low number of COVID-19 cases were reported in suburban districts in the west and south of Hanoi city.

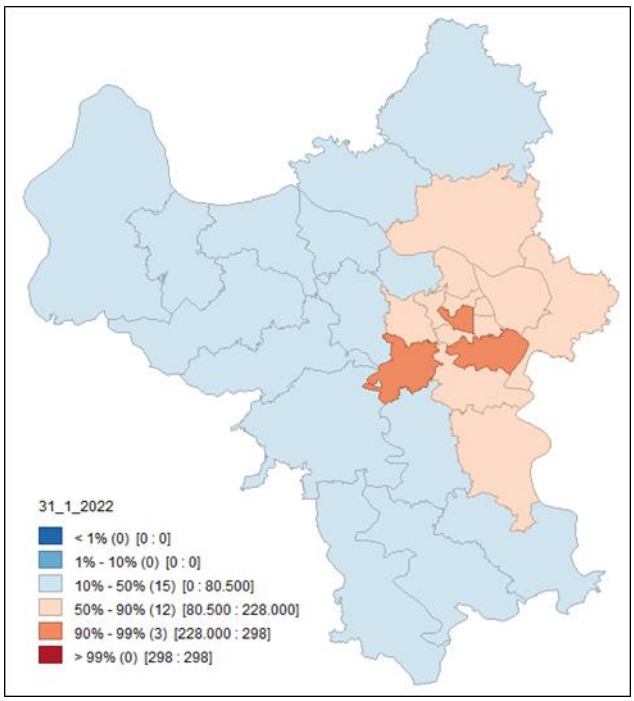

Figure 1 Percentile map of COVID-19 cases confirmed on 31 January 2022 in Hanoi city

3. Methods

Getis-Ord's G_i^* statistic-based hotspot analysis characterizes the presence of hotspots (high clustered values) and coldspots (low clustered values) over an entire area by looking at each feature within the context of its neighboring features (20). It is, therefore, *Getis-Ord's* G_i^* statistic was used to identify the counties of high and low numbers of COVID-19 cases (6,7). The form of *Getis-Ord's* G_i^* statistic is defined as follows (21):

$$G_i^* = \frac{\sum_{j=1}^{N} W_{ij}x_j - \overline{x}\sum_{j=1}^{N} W_{ij}}{S\sqrt{\frac{N\sum_{j=1}^{N}[W_{ij}^2 - (W_{ij})^2]}{N-1}}} \quad \text{..... (1)}$$

$$\text{with } \overline{x} = \frac{1}{N}\sum_{j=1}^{N} x_j \text{ .....} \quad (2)$$

$$\text{and } S = \sqrt{\frac{\sum_{j=1}^{N} x_j^2}{N} - (\overline{x})^2} \text{ .....} \quad (3)$$

where G_i^* is computed for the number of COVID-19 cases at county i; x_i, x_j, $\bar{x}$, and W_{ij} are defined in equation (1); and N is the total number of neighborhood counties as defined in equation (2). W_{ij} can be constructed using the methods of the first order and second of contiguity (shown in Figure 2).

Similar to those obtained from global and local *Moran's I* statistics,the *Getis-Ord's* G_i^* coefficient at county i (G_i^*) also ranges between -1 and +1. If $G_i^* > 0$ and $p(G_i^*) < \alpha$ then there exists a spatial clustering of high-high values (6,7). In this case, these high-high values, so-called a hotspots, reflects the presence of high numbers of COVID-19 cases among county i and its neighborhood counties ($j \in J_i$). Whereas, if $G_i^* < 0$ and $p(G_i^*) < \alpha$ then there exists a spatial clustering of low-low values (6,7). These low-low values are called a coldspots indicating low numbers of COVID-19 cases among county i and its neighborhood counties ($j \in J_i$). Similar to those in the definition of local Moran's I statistic, if the value of G_i^*close zero and $p(G_i^*) < \alpha$ then there will be neither hotspots nor coldspots or random distribution of COVID-19 cases (7).

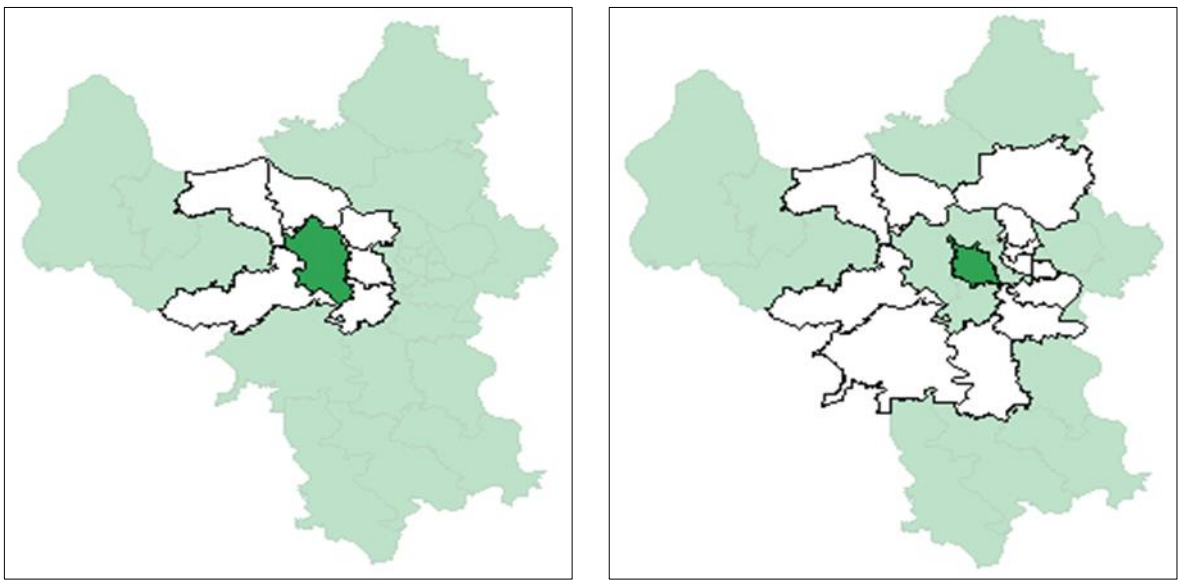

Figure 2 Map of neighbor connectivity using first order (left) and second (right) of contiguity

A lots of attempts have been put on the use of *Getis-Ord's* G_i^* statistic with the help of ArcGIS software using *Getis z-scores* (20,22,23) defined in a study by Mitchel (22). However, as discussed above, the presence of a strongly skewed distribution in the dataset fails the test. It is, therefore, testing for the significance of the *Getis-Ord's* G_i^* statistic in this study was also carried out by a randomization test using 999 permutations. In this work, with the help of the spatial statistics software, GeoDA, developed by (24), a randomization test was used to test the significance of spatial autocorrelation statistics. Spatial autocorrelation statistics were generated and tested at the significance of 0.05 using 999 permutations.

4. Results and discussion

4.1. Analysis of spatial weight matrix

The properties of spatial weight matrix based on the first and second order of contiguity were statistically summarized in Table 1. Data from Table 1 illustrate a total of 30 districts have been used to determine the spatial weight matrix. In both cases, the number of min neighbors count is two. For the method of the first order of contiguity, the max, mean and median of neighbors are 7, 4.8 and 5, respectively. For the second order of contiguity, values of max, mean and median of neighbors are all greater than those obtained from the first case 1 with values of 13, 8.0 and 8.0 respectively (see Table 1).

Table 1 Summary table of properties of spatial weight matrix

Properties	**Values**	
	First order of contiguity	**Second order of contiguity**
Number of observations	30	30
Min neighbors	2	2
Max neighbors	7	13
Mean neighbors	4.8	8.0
Median neighbors	5	8.0
% non-zero	16.0 %	26.7 %

4.2. Analysis of COVID-19 hotspots and coldspots

The results and spatial distribution of COVID-19 hotspot detection in the case of using the first order of contiguity were statistically summarized in Table 2 and was also shown in Figure 3, respectively. Data from the cluster map in Figure 3-a shows a total of 6 hotspots, 6 coldspots and 18 districts with no statistically significance at the 0.05 level. Hotspots were mainly concentrated in the eastern districts of Hanoi. Six COVID-19 hotspots were Cau Giay (116 cases), Thanh Xuan (137 cases), Thanh Tri (143 cases), Gia Lam (166 cases), Long Bien (110 cases), and Hai Ba Trung (127 cases). In addition, Hoang Mai (266 cases), Ha Dong (260 cases) and Dong Da (298) were districts with a large number of infections but no COVID-19 hotspots have been detected in these districts. Surrounding districts having a smaller number of COVID-19 infected cases can account for this issue. In addition, coldspots were detected in the western suburban districts of Hanoi such as Ba Vi (0 cases), Son Tay (0 cases), Thach That (0 cases), Phuc Tho (0 cases), and other suburban districts in the south of Hanoi city such as Ung Hoa (3 cases) and My Duc (0 cases). Data from Figure 3 (left) also shows that there was no existence of hot spots and cold spots in the remained 18 districts at the statistical significance level of 0.05.

Table 2 Summary table of clusters and significant values

Districts	**COVID-19 cases**	**Community types**	**First order of contiguity**			**Second order of contiguity**		
			Getis	**Clusters**	**P value**	**Getis**	**Clusters**	**P value**
Ba Dinh	107	Urban	0.05	0	0.08	0.06	1	0.00
Ba Vi	0	Rural	0.00	2	0.00	0.00	2	0.03
Bac Tu Liem	77	Urban	0.04	0	0.27	0.04	0	0.16
Cau Giay	116	Urban	0.05	1	0.04	0.05	1	0.02
Chuong My	1	Rural	0.02	0	0.23	0.03	0	0.37
Dan Phuong	4	Rural	0.02	0	0.22	0.03	0	0.40
Dong Anh	178	Rural	0.03	0	0.27	0.05	0	0.08
Dong Da	298	Urban	0.06	0	0.16	0.07	1	0.00

Gia Lam	166	Rural	0.07	1	0.02	0.04	0	0.40
Ha Dong	260	Urban	0.04	0	0.46	0.04	0	0.31
Hai Ba Trung	127	Urban	0.07	1	0.01	0.06	1	0.00
Hoai Duc	56	Rural	0.03	0	0.48	0.03	0	0.22
Hoan Kiem	95	Urban	0.06	0	0.06	0.06	1	0.02
Hoang Mai	266	Urban	0.06	0	0.05	0.06	1	0.00
Long Bien	110	Urban	0.05	1	0.04	0.04	0	0.32
Me Linh	10	Rural	0.02	0	0.15	0.02	0	0.13
My Duc	0	Rural	0.00	2	0.05	0.02	0	0.43
Nam Tu Liem	196	Urban	0.05	0	0.10	0.04	0	0.16
Phu Xuyen	4	Rural	0.02	0	0.27	0.03	0	0.39
Phuc Tho	0	Rural	0.00	2	0.00	0.03	0	0.34
Quoc Oai	2	Rural	0.02	0	0.23	0.02	2	0.03
Soc Son	17	Rural	0.03	0	0.43	0.02	0	0.31
Son Tay	0	Town	0.00	2	0.00	0.01	2	0.02
Tay Ho	84	Urban	0.04	0	0.21	0.05	0	0.05
Thach That	0	Rural	0.00	2	0.00	0.02	0	0.26
Thanh Oai	55	Rural	0.03	0	0.49	0.04	0	0.17
Thanh Tri	143	Rural	0.06	1	0.01	0.04	0	0.26
Thanh Xuan	137	Urban	0.07	1	0.00	0.03	0	0.38
Thuong Tin	126	Rural	0.03	0	0.36	0.05	0	0.06
Ung Hoa	3	Rural	0.00	2	0.02	0.04	0	0.18

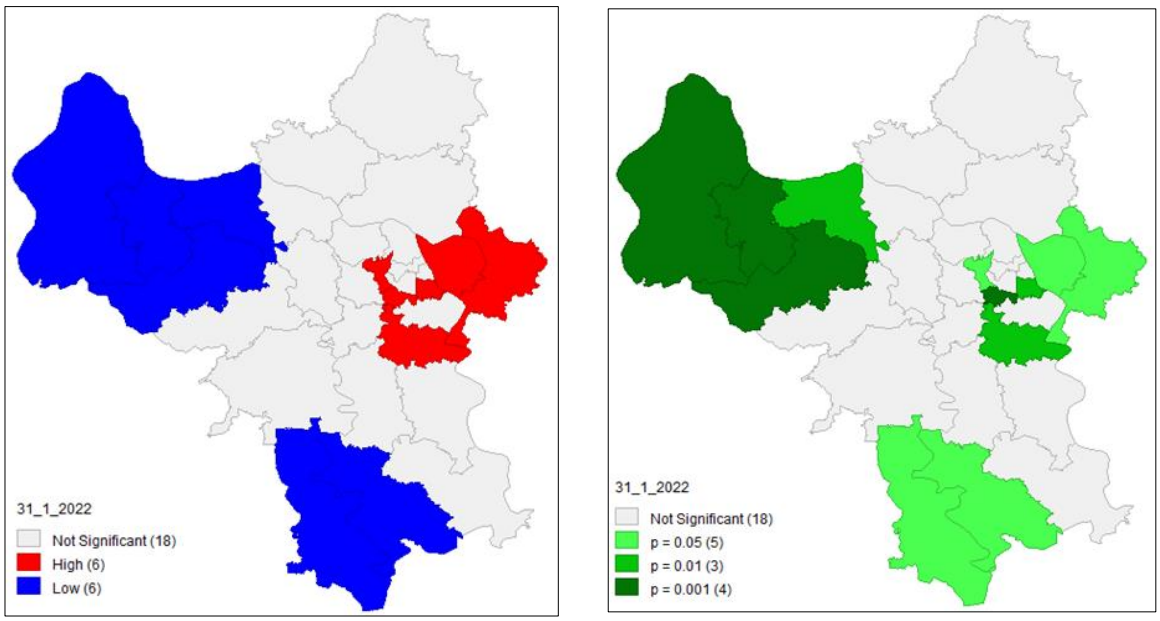

Figure 3 Map of COVID-19 hotspots and coldspots using first order of contiguity: cluster map (left) and significance map (right)

Data from Figure 3 (right) shows the level of statistical significance for each district of Hanoi city. Statistical levels were expressed on 4 scale ranging from statistical unsignificance (p-value > 0.05) to statistical significance at 0.05, 0.01 and 0.001 levels. Data from Figure 3 (right) illustrates that there were 4 clusters with highly significant at 0.001 including Thanh Xuan (137 cases), Thach That (0 cases), Ba Vi (0 cases) and Son Tay (0 cases). Three districts were clustered with high significant at the 0.01 level including Thanh Tri (143 cases), Hai Ba Trung (127 cases) and Phuc Tho (0 cases). Whereas, five clusters were statistically significant at the 0.05 level, including Cau Giay (116 cases), Long Bien (110 cases), Gia Lam (166 cases), Ung Hoa (3 cases) and My Duc (0 cases).

Hotspots of COVID-19 in Hanoicity in the case of using the second order of contiguity were statistically summarized and their spatial distribution was shown in Figure 4. Data from the cluster map in Figure 4 (right) shows that, similar to those obtained from using the first order of contiguity, a total of 6 hotspots were also detected. In addition, three coldspots were identified, and 21 districts with statistically unsignificance at the 0.05 level. The six COVID-19 hotspots were distributed mainly in the eastern districts of Hanoi city. Six hotspots were discovered in urban districts including Cau Giay (116 cases), Dong Da (298 cases), Hoan Kiem (95 cases), Hai Ba Trung (127 cases) and Hoang Mai (266 cases). In addition, Thanh Xuan (137 cases), Thanh Tri (143 cases), Gia Lam (166 cases), Long Bien (110 cases), and Ha Dong (260 cases) were districts having a large number of cases but no COVID-19 hot spots were successfully detected. Similar to those obtained from the above case, three coldspots were also mainly detected in the western suburban districts of Hanoi. However, no COVID-19 clusters were found in the southern districts. The three coldspots include Ba Vi (0 cases), Son Tay (0 cases), and Quoc Oai (2 cases). Some suburban districts had low infection cases but no COVID-19 coldspots were detected in these areas such as Thach That (0 cases) and Phuc Tho (0 cases), Ung Hoa (3 cases) and My Duc (0 cases). Data from Figure 4 (left) also shows that no COVID-19 clusters were found in the remained 21 districts because they were not statistically significant at the 0.05 level.

The data from Figure 4 (right) shows the level of statistical significance for each district of Hanoi city in the case of using the second order of contiguity to construct the spatial weight matrix. The levels of statistical significance were also expressed on 4 scales ranging from statistically unsignificance (> 0.05) to statistical significance at 0.05, 0.01 and 0.001 levels. Data in Figure 4 (right) illustrates that there was only one cluster reached a very high level of significance at the 0.001 level, which was found in Hoang Mai (266 cases). Three COVID-19 clusters were high significant at the 0.01 level including Hai Ba Trung (127 cases), Dong Da (298 cases) and Ba Dinh (107 cases). Five clusters were statistically significant at the 0.05 level including Ba Vi (0 cases), Son Tay (0 cases), Quoc Oai (2 cases), Cau Giay (116 cases) and Hoan Kiem (95 cases).

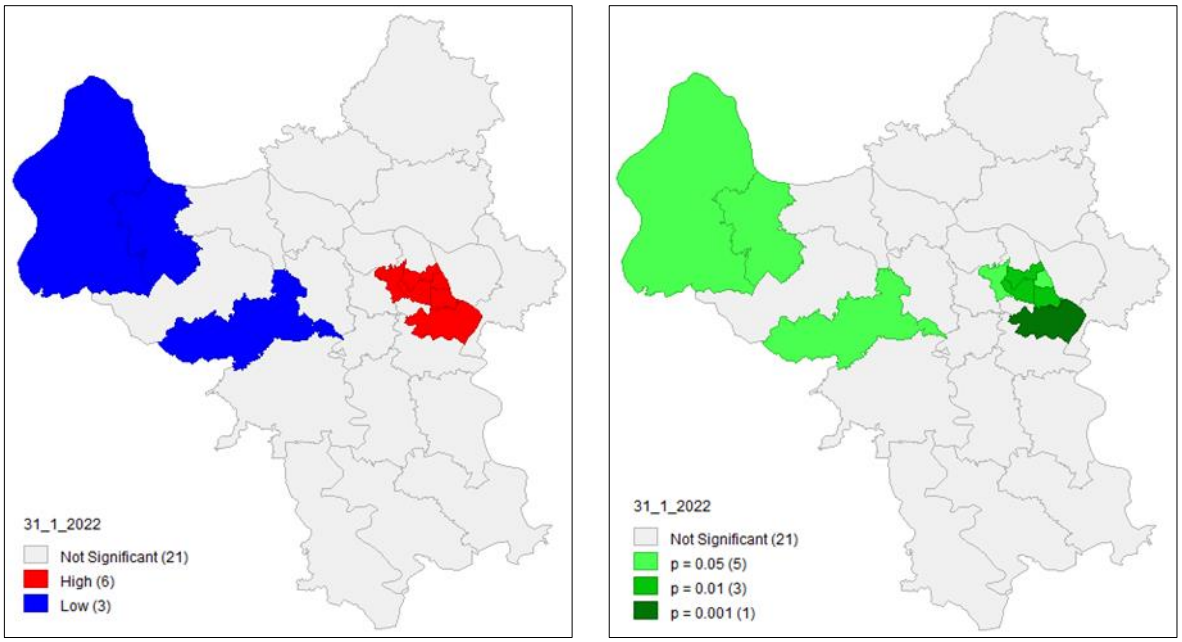

Figure 4 Map of COVID-19 hotspots and coldspots using second order of contiguity: cluster map (left) and significance map (right)

5. Conclusion

In this study, analysis of hotspots and coldspots of COVID-19 was carried out in Hanoi city, Vietnam. The Getis-Ord's G_i^* statistic-based hotspot analysis was employed to detect hotspots and coldspots of COVID-19 pandemic. Two methods for constructing spatial weight matrix, namely the first order and second of contiguity, have been used. It was found from a case study of COVID-19 cases reported on 31 January 2022 in Hanoi city, a total of 6 hotspots and 6 coldspots of COVID-19 were detected statistically significance at the 0.05 level using the first order and second of contiguity. For the case of using the second order of contiguity, six hotspots and three coldspots were successfully identified. Hotspots were mainly concentrated in urbarn districts in the east of Hanoi. Coldspots were detected in the western suburban districts. The study results has proven the effective use of Getis-Ord's statistics to detect hotspots and coldspots of COVID-19 pandemic. Findings in this study greatly contribute to our understanding of the spread of the COVID-19 pandemic. This study also highlighted the need for more studies on COVID-19 hotspot using spatial statistics for COVID-19 prevention and control strategies

Compliance with ethical standards

Acknowledgments

The authors would like to thank Hanoi Center for Diseas Control (Hanoi CDC) for providing the data, and the editors and anonymous reviewers for their insightful and helpful criticism, which served to significantly raise the paper's quality.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Shi Z, Chen H, Fan K, Chen P. Some thoughts and strategies of planning for the impact of "COVID-19" epidemic in Yunnan plateau basin. In: E3S Web of Conferences. EDP Sciences, 2020. p. 3044.

[2] Spagnuolo G, De Vito D, Rengo S, Tatullo M. COVID-19 outbreak: an overview on dentistry. Int J Environ Res Public Health. 2020, 17(6):2094.

[3] WHO. WHO Coronavirus (COVID-19) Dashboard [Internet]. 2023. Available from: https://covid19.who.int/

[4] Carballada AM, Balsa-Barreiro J. Geospatial analysis and mapping strategies for fine-grained and detailed COVID-19 data with GIS. ISPRS Int J Geo-Information. 2021, 10(9):602.

[5] Carroll LN, Au AP, Detwiler LT, Fu T, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. J Biomed Inform. 2014, 51:287–98.

[6] Hoang A, Nguyen T. Identifying Spatio-Temporal Clustering of the COVID-19 Patterns Using Spatial Statistics: Case Studies of Four Waves in Vietnam. Int J Appl Geospatial Res. 2022, 13(1):1–15.

[7] Vu D-T, Nguyen T-T, Hoang A-H. Spatial clustering analysis of the COVID-19 pandemic: A case study of the fourth wave in Vietnam. Geogr Environ Sustain. 2021, 14(4).

[8] Marshall RJ. A review of methods for the statistical analysis of spatial patterns of disease. J R Stat Soc Ser A (Statistics Soc. 1991, 154(3):421–41.

[9] Rosli NM, Shah SA, Mahmood MI. Geographical Information System (GIS) application in tuberculosis spatial clustering studies: a systematic review. Malaysian J Public Heal Med. 2018, 18(1):70–80.

[10] Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect. 2004, 112(9):998–1006.

[11] Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008, 451(7181):990–3.

[12] Rozenfeld Y, Beam J, Maier H, Haggerson W, Boudreau K, Carlson J, et al. A model of disparities: risk factors associated with COVID-19 infection. Int J Equity Health. 2020, 19(1):1–10.

[13] Nykiforuk CIJ, Flaman LM. Geographic information systems (GIS) for health promotion and public health: a review. Health Promot Pract. 2011, 12(1):63–73.

[14] Krieger N. Place, space, and health: GIS and epidemiology. Epidemiology. 2003, 14(4):384–5.

[15] Nazia N, Butt ZA, Bedard ML, Tang W-C, Sehar H, Law J. Methods used in the spatial and spatiotemporal analysis of COVID-19 epidemiology: a systematic review. Int J Environ Res Public Health. 2022, 19(14):8267.

[16] Das A, Ghosh S, Das K, Basu T, Dutta I, Das M. Living environment matters: Unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. Sustain Cities Soc. 2021, 65:102577.

[17] Xie Z, Qin Y, Li Y, Shen W, Zheng Z, Liu S. Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. Sci Total Environ. 2020, 744:140929.

[18] Ramírez-Aldana R, Gomez-Verjan JC, Bello-Chavolla OY. Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. PLoS Negl Trop Dis. 2020, 14(11):e0008875.

[19] Castro MC, Kim S, Barberia L, Ribeiro AF, Gurzenda S, Ribeiro KB, et al. Spatiotemporal pattern of COVID-19 spread in Brazil. Science (80-). 2021, 372(6544):821–6.

[20] Nguyen TT, Vu TD. Use of hot spot analysis to detect underground coal fires from landsat-8 TIRS data: A case study in the Khanh Hoa coal field, North-East of Vietnam. Environ Nat Resour J. 2019, 17(3).

[21] Cliff AD, Ord JK. Spatial processes: models & applications. (No Title). 1981,

[22] Mitchell A. The ESRI guide to GIS analysis: geographic patterns & relationships. Vol. 1. ESRI, Inc., 1999.

[23] Alves JD, Abade AS, Peres WP, Borges JE, Santos SM, Scholze AR. Impact of COVID-19 on the indigenous population of Brazil: A geo-epidemiological study. Epidemiol Infect. 2021, 149:e185.

[24] Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. In: Handbook of applied spatial analysis: Software tools, methods and applications. Springer, 2009. p. 73–89.