RESEARCH PAPER

SPATIAL CLUSTERING ANALYSIS OF THE COVID-19 PANDEMIC: A CASE STUDY OF THE FOURTH WAVE IN VIETNAM

Danh-tuyen Vu[1], Tien-thanh Nguyen[1*], Anh-huy Hoang[3]

[1]Faculty of Surveying, Mapping and Geographic Information, Hanoi University of Natural Resources and Environment, No. 41A, PhuDien Road, North-TuLiem District, Hanoi,100000, Vietnam

[2]Faculty of Environment, Hanoi University of Natural Resources and Environment, No. 41A, PhuDien Road, North-TuLiem District, Hanoi, 100000, Vietnam

***Corresponding author:** tdgis_ntthanh@163.com; ntthanh@hunre.edu.vn

Received: July 31th, 2021 / Accepted: November 9th, 2021 / Published: December 31st, 2021

https://DOI-10.24057/2071-9388-2021-086

ABSTRACT. An outbreak of the 2019 Novel Coronavirus Disease (COVID-19) in China caused by the emergence of Severe Acute Respiratory Syndrome CoronaVirus 2 (SARSCoV2) spreads rapidly across the world and has negatively affected almost all countries including such the developing country as Vietnam. This study aimed to analyze the spatial clustering of the COVID-19 pandemic using spatial auto-correlation analysis. The spatial clustering including spatial clusters (high-high and low-low), spatial outliers (low-high and high-low), and hotspots of the COVID-19 pandemic were explored using the local Moran's I and *Getis-Ord's* G_i^* statistics. The local Moran's I and Moran scatterplot were first employed to identify spatial clusters and spatial outliers of COVID-19. The *Getis-Ord's* G_i^* statistic was then used to detect hotspots of COVID-19. The method has been illustrated using a dataset of 86,277 locally transmitted cases confirmed in two phases of the fourth COVID-19 wave in Vietnam. It was shown that significant low-high spatial outliers and hotspots of COVID-19 were first detected in the North-Eastern region in the first phase, whereas, high-high clusters and low-high outliers and hotspots were then detected in the Southern region of Vietnam. The present findings confirm the effectiveness of spatial auto-correlation in the fight against the COVID-19 pandemic, especially in the study of spatial clustering of COVID-19. The insights gained from this study may be of assistance to mitigate the health, economic, environmental, and social impacts of the COVID-19 pandemic.

KEYWORDS: Spatial clustering; Spatial auto-correlation; COVID-19 pandemic; Vietnam's fourth wave

CITATION: Danh-tuyen Vu, Tien-thanh Nguyen, Anh-huy Hoang (2021). Spatial Clustering Analysis of the COVID-19 Pandemic: A Case Study of the Fourth Wave in Vietnam. Geography, Environment, Sustainability, Vol.14, No 4, p. 140-147 https://DOI-10.24057/2071-9388-2021-086

ACKNOWLEDGEMENTS: We wish to thank the editors and three anonymous reviewers for their valuable and constructive comments and suggestions on this paper that have helped us to greatly improve the quality of the paper.

Conflict of interests: The authors reported no potential conflict of interest.

INTRODUCTION

The 2019 COVID-19 is a pandemic illness that was discovered in Wuhan of China at the end of 2019. The COVID-19 epidemic quickly spreads worldwide rapidly to emerge as a global public health concern (Das et al. 2021). Globally, as of 26 October 2021, a total of more than 243.6 million confirmed cases of the COVID-19 and more than 4.9 million deaths were reported (WHO 2021). The COVID-19 pandemic has been described as a social, human, and economic crisis (United Nations 2020). It is, therefore, the understanding of the spatial distribution of the COVID pandemic, in general, and of the spatial clustering, in particular, plays an important role in the fight of COVID-19.

In studies of the COVID-19 pandemic, Shepherd (2020) indicated that geography is considered a key part of fighting the COVID-19 coronavirus outbreak. Especially, after Rose-Redwood et al. (2020) discovered the COVID-19 pandemic is thoroughly spatial in nature, a lot of efforts have been made on the study of the COVID-19 pandemic from a geographical perspective to better understand the spatial distribution and better manage the COVID-19 infection. Using *Getis-Ord's* G_i^* statistic and geographically weighted principal component analysis, Das et al. (2021) successfully examined the impact of living environment deprivation on the COVID-19 hotspot in Kolkata megacity, India. The result study of Das et al. (2021) revealed that living environment deprivation was an important determinant of spatial clustering of COVID-19 hotspots in the Kolkata megacity. Xie et al. (2020) used the exploratory spatial data analysis and the geodetector method 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. The results of Xie et al. (2020) demonstrate two things. First, the population inflow from Wuhan and the strength of economic connection are the main factors affecting the epidemic's spread. Second, when the average temperature in winter is maintained at 11-16°C, the epidemic spread rate is higher. Later, with the aim to analyze the spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with the

environmental factors, based on the incidences of new local COVID-19 cases in Beijing, Han et al. (2021) investigated the spatial clustering characteristics of the COVID-19 pandemic in Beijing using spatial autocorrelation analysis. In line with those obtained by Xie et al. (2020), Han et al. (2021) also indicated that population density and distance to the market are key factors of the pandemic. When a rapid increase in the number of COVID-19 cases was reported in Iran in 2020, Ramírez-Aldana et al. (2020) developed a spatial statistical approach to describe how COVID-19 cases are spatially distributed and to identify significant spatial clusters of cases and how socioeconomic and climatic features of Iranian provinces might predict the number of cases. In that study, Ramírez-Aldana et al. (2020) successfully applied global (Moran's I) and local indicators of spatial autocorrelation (LISA), both univariate and bivariate, to derive significant clustering of COVID-19 pandemic. In South Korea, to understand the COVID-19 clustering across districts in South Korea and how the spatial pattern of COVID-19 changes, Kim and Castro (2020) successfully applied the global Moran's I statistic and the retrospective space-time scan statistic to analyze spatio-temporal clusters of COVID-19. A similar conclusion was reached by Choi (2020), Kim and Castro (2020) also showed that the spatial pattern of clusters changed and the duration of clusters became shorter over time in this country. Most recently, when identifying spatial patterns of COVID-19 disease clustering in India using another global spatial autocorrelation statistic, the *Getis-Ord's* G^*_i statistic, Bhunia et al. (2021) discovered that this statistic can help public health professionals to effectively identify risk areas for disease and take decisions in real-time to control this viral disease. Similar to those obtained from these studies, a recent study by Robinson (2000) has also indicated that commonly used statistics such as global spatial statistics (Moran's I, *Getis-Ord's* G^*_i and Geary's C) and LISA have been successfully applied in epidemiological studies in general and in the study of COVID-19 pandemic in particular. Among these spatial statistics, local Moran's I and *Getis-Ord's* G^*_i statistics have been most widely used for the analysis of spatial auto-correlation (Zhang and Zhang 2007) because of their effectiveness in displaying the spatial distribution of infectious diseases (Tran et al. 2004; Wang et al. 2015), thus, these two statistics will be employed to analyze the spatial clustering of the COVID-19 pandemic in this study.

A Delta-driven fourth wave of COVID-19 has profoundly affected the world including such countries in Southeast Asia as Vietnam. Although, a lot of efforts have been put into the study of the COVID-19 pandemic in Vietnam (Duy et al. 2021; Hoang et al. 2020; Huang et al. 2020; La et al. 2020; Le and Tran 2021; Nguyen and Vu 2020), so far, very little attention has been paid to the role of geographical methods. Most recently, Hoang et al. (2020) described the pattern of the COVID-19 epidemic in Vietnam, however, no spatial autocorrelation also has been taken into account. Particularly, to date, no previous study has investigated the spatial clustering of the COVID-19 pandemic in Vietnam. Accordingly, this study aims to analyze spatial clustering of the COVID-19 pandemic using spatial auto-correlation analysis. It is believed that this is the first study on the identification of spatial clustering of COVID-19 pandemic in Vietnam that accounts for spatial auto-correlation among locally transmitted cases COVID-19 cases.

STUDY AREA AND DATA USED

According to the Ministry of Health of Vietnam (MHV), as of 24 June 2021, the COVID-19 pandemic in Vietnam can be divided into four COVID-19 waves and resulted in a total of 89,992 confirmed cases (87,847 locally transmitted cases and 2,145 internationally imported cases) and 63 deaths (MHV 2021). The first wave started from 23 January to 24 July 2020 causing 415 confirmed cases (106 locally transmitted cases and 309 internationally imported cases). Vietnam was ranked 94th on the list of 206 countries and territories affected by COVID-19 (Ha et al. 2020). The second wave ranged from 25 July 2020 to 27 January 2021 with 1,136 confirmed cases (554 locally transmitted cases and 582 internationally imported cases) (Ha et al. 2020) and 35 deaths as of February 2021 (Phuong et al. 2021). The third wave was from 28 January 2021 to 26 April 2021 with 1,301 cases (910 locally transmitted cases and 391 internationally imported cases) (MHV 2021). A total of 25 COVID-19 clusters related to the outbreaks had been identified mainly in Van Don international airport (VnExpress 2021a) and Dong Trieu town in Quang Ninh province (VnExpress 2021b), and POYUN company and Haiduong city in Hai Duong province (HanoiTimes 2021). The latest data show that, as of 24 June 2021, the on-going fourth wave has resulted in 87,140 (86,277 locally transmitted cases and 863 internationally imported cases) since a series of COVID-19 confirmed cases occurred due to viral virus strain from the UK and India was declared on 27 April 2021 in industrial zones in Bac Giang and Bac Ninh provinces, North-East of Vietnam (VietNamNews 2021a; VnExpress 2021c) and the presence of the Delta variant (VietNamNews 2021b). In this study, the focus will be made on the fourth wave of COVID-19 in all 63 provinces and cities in Vietnam. Consequently, a total of 86,277 locally transmitted cases confirmed in the fourth wave of COVID-19 was used to identify the spatial clustering of the COVID-19 pandemic. Data from Figure 1 demonstrates the daily number of new locally transmitted cases were reported in Vietnam by MHV (2021) since the ongoing fourth wave of COVID-19 infections started on 27 April 2021. The spatial distribution of these locally transmitted cases in the first, second phase of the fourth COVID-19 wave and the whole fourth wave was shown in Figures 2-a, b, and c, respectively.

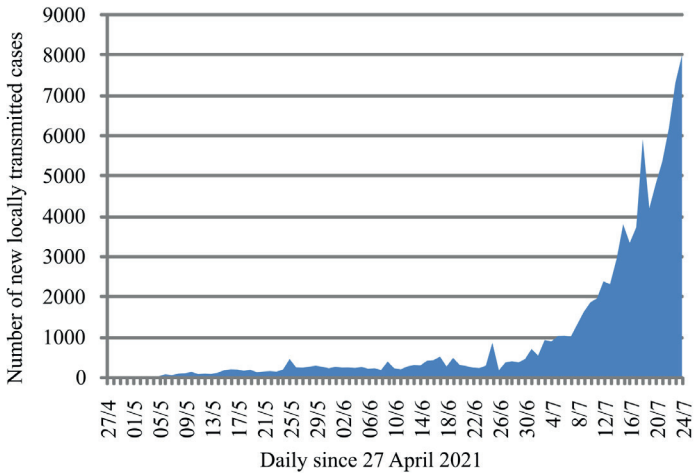

Fig. 1. Daily new confirmed cases in the fourth wave of COVID-19 in Vietnam

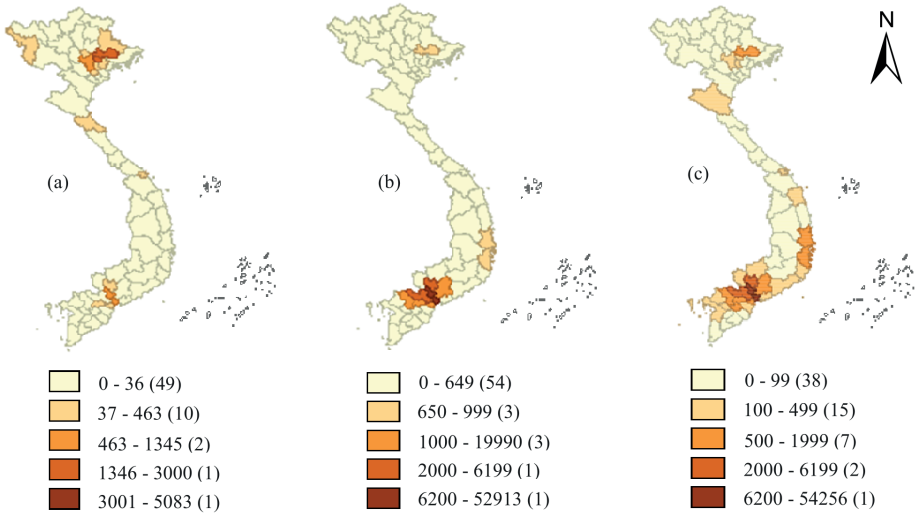

Fig. 2. Spatial distribution of confirmed cases in the fourth wave of COVID-19 in Vietnam

METHODOLOGY

The workflow for spatial clustering analysis of the COVID-19 pandemic is shown in Figure 3. The number of locally transmitted COVID-19 cases was firstly collected from a database provided by the official websites of the Ministry of Health of Vietnam (MHV 2021). After the spatial weight matrix was constructed, with the help of GeoDA software developed by Anselin et al. (2016), Moran's I and *Getis-Ord's* G_i^* statistics were then computed to create LISA and hotspot maps. These maps were finally used for the evaluation of the spatial clustering of the COVID-19 pandemic.

Identifying spatial clustering of the COVID-19 pandemic using *Moran's I* statistic

In this study, the degree of spatial clustering of the COVID-19 pandemic as a whole is measured using global *Moran's I* statistic (Cliff and Ord 1981; Getis and Ord 1996). The global Moran's I statistic is defined as follow:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad (1)$$

where x_i and x_j are the cumulative numbers of confirmed cases for province/city *i* and province/city *j* in the fourth wave of COVID-19; $\bar{x}$ is the mean of COVID-19 cases and be given $x = \sum_{i=1}^{n} \frac{x_i}{n}$; n is the total number of provinces/cities (all 63 provinces and cities in Vietnam) in the whole study area; and W_{ij} is a (n x n) spatial queen contiguity weight matrix.

The global *Moran's I* coefficient takes values in the interval [-1, +1]. Positive spatial autocorrelation in the data translates into positive values of *Moran's I*, whereas negative spatial autocorrelation produces negative values of *Moran's I* (Nguyen and Vu 2019a). The values of global *Moran's I* coefficients are close to zero indicating no spatial clustering or random distribution of COVID-19 pandemic.

The global *Moran's I* reflects the presence or lack of spatial clustering as a whole (Nguyen et al. 2016; Nguyen and Vu 2019a). Therefore, the spatial clustering of low and high COVID-19 pandemic at each province/city was then measured utilizing the local *Moran's I* statistic. The local *Moran's I* statistic (I_i) for COVID-19 pandemic at province/city i is given by the following equation (Anselin 1995):

$$I_i = \frac{\left(x_i - x\right)}{\sigma^2} \sum_{j\#i,\, j\in J_i}^{N} W_{ij}\left(x_j - \bar{x}\right) \quad (2)$$

where x_i, x_j, $\bar{x}$, and W_{ij} are defined in equation (1); *N* is the total number of neighborhood provinces/cities; J_i denotes the neighborhood set of COVID-19 confirmed cases at province/city *i*; j#i implies that the sum of all $(x_j - \bar{x})$ of nearby neighbourhood provinces/cities of province/city *i* but not including x_j; and σ^2 is the variance of *x*, given in equation (3).

$$\sigma^2 = \frac{1}{N}\sum_{j=1}^{N}\left(x_i - \bar{x}\right) \quad (3)$$

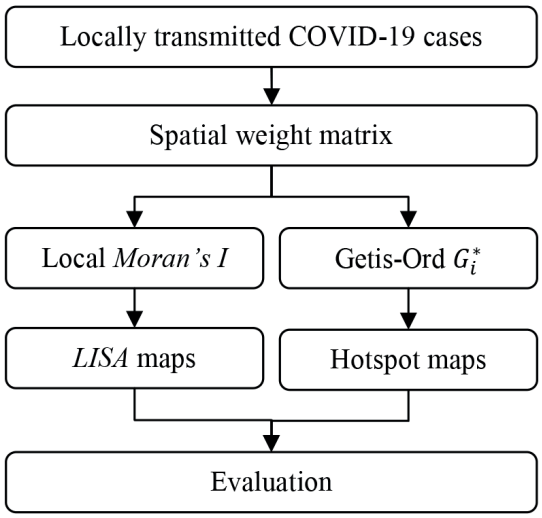

Fig. 3. The workflow for spatial clustering analysis of the COVID-19 pandemic

Local *Moran's I* reflects the degree of spatial clustering of the COVID-19 pandemic at each province/city. Similar to the global *Moran's I* statistic, the local *Moran's I* value at province/city i (I_i) also ranges between -1 and +1. If the local *Moran's I* coefficient at province/city i equals zero ($I_i = 0$) then there is no spatial clustering of COVID-19 cases. If $I_i > 0$ then there will be a positive spatial clustering of COVID-19 cases. If $I_i < 0$ then there will be a negative spatial clustering of COVID-19 cases. A high positive I_i shows the province/city i has a similarly high or low number of COVID-19 cases as its neighbors and is called the «spatial cluster» by Nguyen and Vu (2019a). In this case, when there exists a positive local spatial autocorrelation, the local *Moran's I* statistic identifies two types of spatial clusters for COVID-19 cases: high-high clusters and low-low clusters. Based on the findings in previous studies (Nguyen et al. 2016; Nguyen and Vu 2019a), if the p-value of local Moran's I at location *i*, $p(I_i)$, is less than a predetermined level (*a*), $p(I_i)<a$, $I_i>0$ *and* $x_i-\bar{x}>0$, then x_i and $x_{j\in Ji}$ belong to a spatial cluster between high numbers of COVID-19 cases. This means a high number of COVID-19 at province/city *i* are clustered with high numbers of COVID-19 at neighborhood provinces/cities. If $p(I_i)<a$, $I_i>0$, *and* $x_i-\bar{x}<0$, then x_i and $x_{j\in Ji}$ belong to a spatial cluster between the low number of COVID-19 cases. This means a low number of COVID-19 cases at province/city *i* is clustered with low numbers of COVID-19 at neighborhood provinces/cities. Meanwhile, Lalor and Zhang (2001) defined spatial outliers as those values that are significantly different from the values of their surrounding locations. In this case, similar to those reported in Nguyen et al. (2016) when there exists a negative local spatial autocorrelation, local *Moran's I* identifies two types of spatial outliers: low-high and high-low clusters. If $p(I_i)<a$, $I_i<0$, *and* $x_i-\bar{x}>0$ then a province/city with a high number of COVID-19, x_i, is surrounded by provinces/cities with low numbers of COVID-19, $x_{j\in Ji}$, (high-low outliers). If $p(I_i)<a$, $I_i<0$, *and* $x_i-\bar{x}>0$ then a province/city with a low number of COVID-19, x_i, is surrounded by provinces/cities with high numbers of COVID-19, $x_{j\in Ji}$, (low-high outliers).

Anselin (1995) indicates that testing for the significance of spatial autocorrelation statistics such as the global and local Moran's I, and *Getis-Ord's* G^*_i can be carried out based on an assumption of a normal distribution. However, these statistics are very sensitive to a strongly skewed distribution (Hoang et al. 2017; Nguyen et al. 2014; Nguyen 2018; Nguyen et al. 2016; Nguyen and Vu 2019a; Nguyen and Vu 2019b) due to the existence of a high and very high number of COVID-19 cases in some provinces or cities. Wherefore, in this study, testing for the significance of these spatial autocorrelation statistics was carried out by a randomization test which recalculates the statistic many times to generate a reference distribution (Anselin 2005). The data values are reassigned among the N fixed locations in the randomization test, providing a randomization distribution against which we can judge our observed value (Nguyen 2018). With the help of the spatial statistics software, GeoDA 095i, developed by Anselin et al. (2016), all spatial autocorrelation coefficients including the *Getis-Ord's* G^*_i statistic (see below for more details) were calculated and tested at the significance of 0.05 using 999 permutations.

Identifying hotspots of the COVID-19 pandemic using Getis-Ord's G^*_i statistic

The *Getis-Ord's* G^*_i statistic can be employed to measure the degree of spatial clustering of COVID-19 incidence in a study location (Liu et al. 2021). *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 (Mitchel 2005; Nguyen and Vu 2019b). Accordingly, in this study, *Getis-Ord's* G^*_i statistic was applied to measure the degree of hotspot of COVID-19. Ord and Getis (1995) defined the form of *Getis-Ord's* G^*_i statistic as follows:

$$G_i^* = \frac{\sum_{j=1}^{N} W_{ij}x_j - \bar{x}\sum_{j=1}^{N} W_{ij}}{S\sqrt{\frac{N\sum_{j=1}^{N}\left[W_{ij}^2 - \left(W_{ij}\right)^2\right]}{N-1}}} \quad (4)$$

$$with\ \bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j\ and\ S = \sqrt{\frac{\sum_{j=1}^{N} x_j^2}{N} - \left(\bar{x}\right)^2}$$

where G^*_i is computed for the number of COVID-19 cases at province/city *i*; x_j, x_i, $\bar{x}$, and W_{ij} are defined in equation (1); and *N* is the total number of neighborhood provinces/cities as defined in equation (2).

Similar to those obtained from global and local *Moran's I* statistics, the *Getis-Ord's* G^*_i coefficient at province/city *i* (G^*_i) also ranges from -1 to +1. If $G^*_i > 0$ and $p(G^*_i) < a$ then there exists a spatial clustering of the high number of COVID-19 cases. In this case, these high-high values, so-called a hotspot, reflects the presence of high numbers of COVID-19 cases among province/city i and its neighborhood provinces/cities ($j \in J_i$). On the contrary, if $G^*_i < 0$ and $p(G^*_i) < a$ then there exists a spatial clustering of low-low values. These low-low values are called a coldspot indicating low numbers of COVID-19 cases among provinces/cities i and its neighborhood provinces/cities ($j \in J_i$). Similar to those in the definition of local Moran's I statistic, if the value of G^*_i close zero and $p(G^*_i) < a$ then there will be no spatial clustering (neither hotspots nor coldspots) or random distribution of COVID-19 cases.

Several studies (Alves et al. 2021; Liu et al. 2021; Nguyen and Vu 2019b) have computed the *Getis-Ord's* G^*_i statistic with the help of ArcGIS software using *Getis z-scores* defined in a study by Mitchel (2005). If provinces/cities with 1.65<*Getis z-scores*<1.96, 1.96<*Getis z-scores*<2.58, and *Getis z-scores*>2.58 were considered to be significant at the 90% confidence level ($p < 0.1$), 95% confidence level ($p < 0.05$), and 99% confidence level ($p < 0.01$), respectively. However, as discussed in the previous section on *Moran's I* statistic, the presence of a strongly skewed distribution in the dataset fails the test. 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.

RESULTS AND DISCUSSIONS

Spatial clustering of the COVID-19 pandemic

Global Moran's I coefficients ($p<0.001$) were 0.1 for the first phase, 0.06 for the second phase, 0.05 for the whole of the ongoing fourth COVID-19 wave, respectively (Figure 4). The global *Moran's I* value of 0.1 indicates the strongest spatial association for the first phrase. Whereas, global *Moran's I* values reduce to 0.06 and 0.05 indicating weaker spatial associations for the second phase and the whole of the fourth COVID-19 wave, respectively. However, these global *Moran's I* coefficients were close to zero indicating, as a whole, there were no spatial auto-correlation or random distribution of the COVID-19 pandemic in this fourth wave. In addition, global *Moran's I* ignore the presence of spatial auto-correlation of COVID-19 confirmed cases at local or provincial scales. To overcome this limitation, local

Moran's I-based LISA was employed to measure the degree of spatial clustering at local scales (province/city level).

A total of 9,192 locally transmitted cases were reported in the first phase of the fourth wave of the COVID-19 pandemic in Vietnam. The spatial distribution of the COVID-19 clustering area was illustrated by LISA maps in Figure 5. Data from Figure 5-a demonstrates that, in the first phase of the ongoing COVID-19 wave, the local *Moran's I* statistic identified a total of three high-high clusters in Bac Giang (5,083 cases), Bac Ninh (1,407 cases), and Hanoi (464 cases); and nine low-low clusters including Ninh Thuan (12 cases), Binh Thuan (11 cases), Dak Lak (6 cases) in South-central region and very low number of COVID-19 cases in Southern provinces of Vietnam such as Lam Dong, An Giang, Kien Giang, Can Tho, Hau Giang, and Bac Lieu. Spatial clustering of COVID-19 pandemic including high-high clusters and low-high outliers was mainly identified in the North-Eastern provinces of Vietnam. These spatial clusters and outliers were mainly caused by more than 6,500 infected workers in industrial parks in Bac Giang and Bac Ninh provinces (VietNamNews 2021a; VnExpress 2021c) after a series of COVID-19 confirmed cases declared on 27 April 2021 due to the rapid spread of the viral virus strain from the UK and India (VietNamNews 2021a; VnExpress 2021c).

Data from Figures 5-b and 5-c illustrates that comparing with those obtained in the first phase, the locations of spatial clustering of COVID-19 pandemic have been quickly changed from the North-Eastern region in the first phase to the Southern region of Vietnam in the second phase of the fourth COVID-19 wave (Figure 5-b). The local *Moran's I* statistic successfully identified five high-high, eleven low-low clusters, and two low-high outliers. Five high-high clusters included Ho Chi Minh city (52,913 cases), the epicenter of the ongoing fourth wave, followed by the Southern provinces of Binh Duong (6,146 cases), Long An (2,178 cases) and Dong Nai (1,778 cases), and Tieng Giang (1,245 cases). Two low-high outliers were detected in the provinces of Ba Ria – Vung Tau (471 cases) and Tay Ninh (204 cases). Whereas, eleven low-low clusters were identified in North-Western provinces of Vietnam with a low number of COVID-19 cases. Policy solutions for COVID-19 response used by local authorities in these provinces have shown effectiveness in stopping the community transmission in these areas. Since the presence of the Delta variant was detected in the second phase of the fourth wave (VietNamNews 2021b), Vietnam recorded 77,085 locally transmitted cases mainly detected in Southern provinces. Similar to those obtained in the second phase, spatial clustering of the COVID-19 pandemic in the whole of the fourth COVID-19 wave was mainly detected in cities/provinces in the Southern region of Vietnam. This is due to the number of locally transmitted cases in this phase accounts for 97% of the total number confirmed in the fourth wave of COVID-19. To effectively fight the COVID-19 in the ongoing fourth wave, social distancing rules under the Government's Directive No. 16 are being imposed in these Southern localities of COVID-19 hotspots (VGP 2021). The rapid spread of infections has prompted strict movement restrictions in around one-third of the country, with both Hanoi and Ho Chi Minh City under lockdown (Reuters 2021).

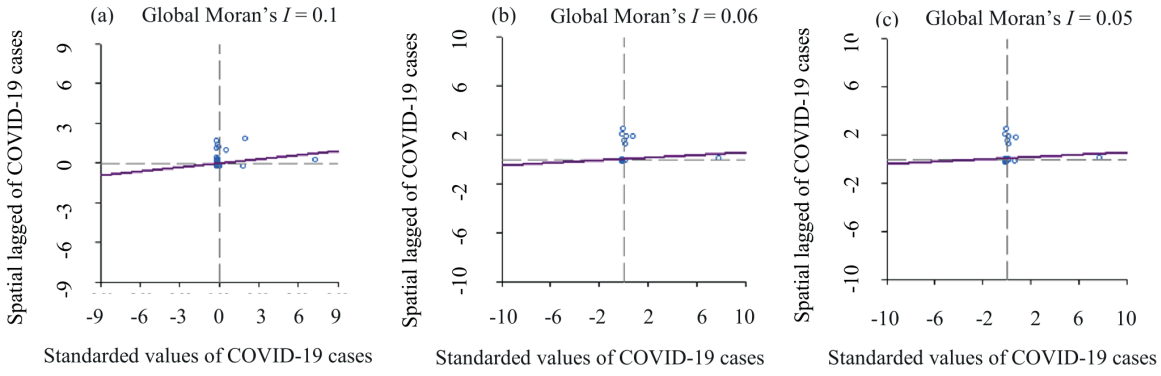

Fig. 4. Moran scatterplot of confirmed cases in the fourth wave of COVID-19 in Vietnam

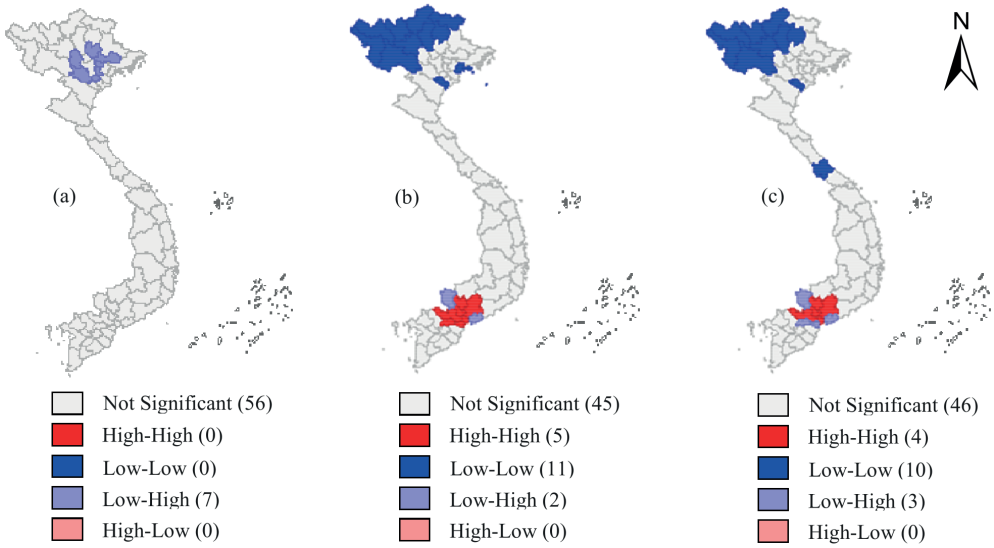

Fig. 5. Spatial clustering of the COVID-19 pandemic in the fourth wave in Vietnam

Hotspots of the COVID-19 pandemic

Maps from Figure 6 illustrate the spatial distribution of hotspots and coldspots of the COVID-19 pandemic in the fourth wave in Vietnam. In the first phase of the fourth COVID-19 wave (Figure 6-a), the local *Getis-Ord's* G_i^* statistic successfully identified a total of six COVID-19 hotspots in the North-Eastern region of Vietnam, including Bac Giang (5,083 cases), Bac Ninh (1,470 cases), Hanoi (464 cases), Hai Duong (51 cases), Thai Nguyen (7 cases), and Quang Ninh (1 case). Especially, two provinces have reported the highest number of new cases including Bac Giang (5,083 cases) and Bac Ninh (1,407 cases). In addition, 11 coldspots with a very low number of COVID-19 cases in provinces and cities of the central and Southern region were successfully detected. A recent study by Nguyen et al. (2021) has shown that effective policy solutions for COVID-19 response of the Vietnamese government such as swift governmental action, strict border control measures, effective communication of health promotion measures, widespread community engagement, expanded testing capacity and effective social measures can be account for the presence of these coldspots in this phrase. But in late April 2021, the highly transmissible Delta variant began to take hold in the Southern region of Vietnam, especially in Ho Chi Minh city – the country's economic engine (Tough 2021) and has caused the second phrase of COVID-19 in Vietnam.

Similar to those obtained using the local Moran's I (see Figure 5-b), the locations of hotspots of the COVID-19 pandemic quickly changed from the North-Eastern region in the first phase to the Southern region of Vietnam in the second phase of the fourth COVID-19 wave (Figure 6-b). In this second phase, the local *Getis-Ord's* G_i^* statistic successfully identified seven COVID-19 hotspots in the southern region of Vietnam and 11 coldspots in the North-Western cities/provinces in Vietnam. Similar to those obtained in the second phase by the local *Moran's I*, seven hotspots of COVID-19 included Ho Chi Minh city (52,913 cases), Binh Duong (6,146 cases), Long An (2,178 cases), and Dong Nai (1,778 cases), Tien Giang (1,245 cases), Ba Ria – Vung Tau (471 cases) and Tay Ninh (204 cases). Similar to those obtained in the second phase, hotspots of the COVID-19 pandemic in the whole of the fourth COVID-19 wave were also mainly detected in cities/provinces in the southern region of Vietnam (Figure 6-c). Among these hotspots of the COVID-19 pandemic, an unknown source of a COVID-19 cluster from a Christian congregation based in Go Vap district caused Ho Chi Minh city the most-affected local with 52,913 confirmed cases and 356 deaths (VnExpress 2021). A most recent study by Tough (2021) has also indicated that social distancing measures used to control previous variants have proven ineffective against the virulent Delta strain in these COVID-19 hotspot areas in Vietnam, and this prompted the Vietnamese authorities to impose increasingly strict lockdowns in these hotspot areas. Another important reason might be the very low testing capacities. In addition, a total of 10 coldspots caused by a very low number of COVID-19 cases was identified in the North-Western provinces. Travel restrictions and the effectiveness of local policies for COVID-19 response play an important role in the effective prevention and control of COVID-19 in these coldspot areas.

CONCLUSIONS

In this study, the spatial clustering of the COVID-19 pandemic was analyzed using spatial auto-correlation analysis. The spatial clustering including spatial clusters (high-high and low-low), spatial outliers (low-high and high-low), and hotspots of the COVID-19 pandemic were explored using the local *Moran's I* and *Getis-Ord's* G_i^* statistics. Results from a case study on 86,277 locally transmitted cases confirmed in two phases of the fourth COVID-19 wave in Vietnam showed that significant low-high spatial outliers and hotspots of COVID-19 were first detected in the North-Eastern region in the first phase. Significant high-high clusters and low-high outliers and hotspots were then identified in the Southern region of Vietnam in the second phase. It can be concluded that spatial statistics (Moran's I and *Getis-Ord's* G_i^* statistics) are of great help in understanding the spatial clustering of the COVID-19 pandemic. The study is limited by the lack of the relative values of the total COVID-19 confirmed cases and deaths to the total population of a municipality in the data used. It is, therefore, further studies regarding the role of the relative values would be worthwhile. Notwithstanding these limitations, the present findings confirm the effectiveness of spatial statistics, particularly spatial clustering analysis, in the fight against the COVID-19 pandemic. These empirical findings not only provide a new understanding of how the COVID-19 pandemic is spatially clustered, but also may be help mitigate the effects of the COVID-19 clustering area and making appropriate decisions related to the health, economic, and social system.

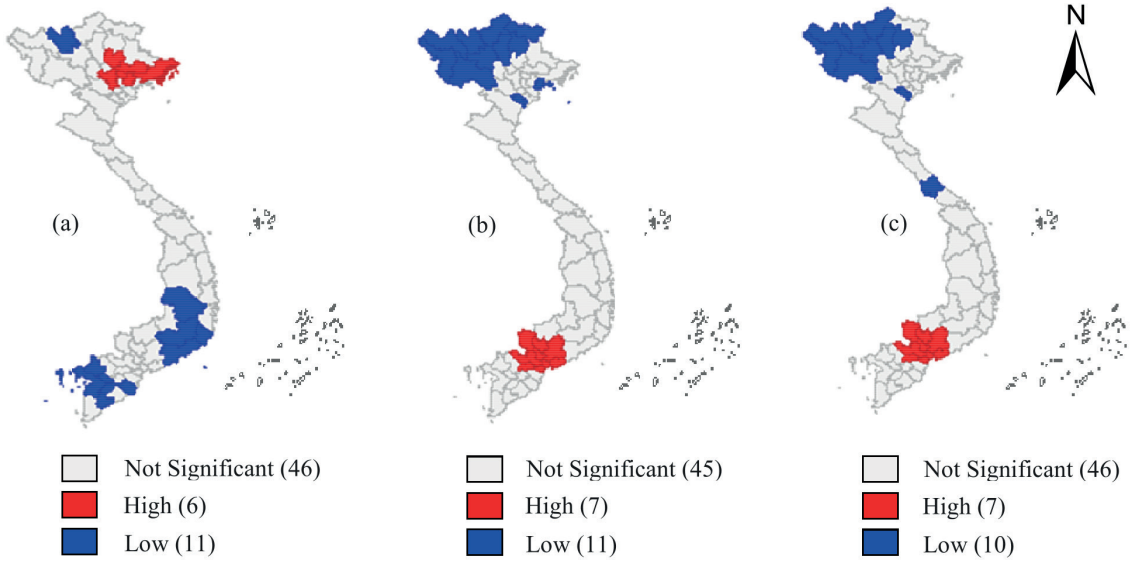

Fig. 6. Hot spots of the COVID-19 pandemic in the fourth wave in Vietnam

AUTHOR CONTRIBUTIONS

Tien-thanh Nguyen conceived and designed the study. Danh-tuyen Vu and Anh-huy Hoang collected the data. Tien-thanh Nguyen and Anh-huy Hoang performed the statistical analysis. Tien-thanh Nguyen and Danh-tuyen Vu prepared the manuscript. All authors discussed the results, implications and commented on the manuscript at all stages.

REFERENCES

Alves J. D., Abade A.S., Peres W.P., Borges J.E., Santos S.M., Scholze A.R. (2021). Impact of COVID-19 on the indigenous population of Brazil: A geo-epidemiological study. MedRxiv, DOI: 10.1101/2021.01.12.21249703.

Anselin L. (1995). Local indicators of spatial association—LISA. Geographical analysis, 27, 93-115, DOI: 10.1111/j.1538-4632.1995.tb00338.x.

Anselin L. (2005). Exploring spatial data with GeoDaTM: a workbook. Center for spatially integrated social science.

Anselin L., Syabri I., Kho Y. (2016). GeoDa: An introduction to spatial data analysis. 2005 Reference Source, DOI: 10.1007/978-3-642-03647-7_5.

Bhunia G.S., Roy S., Shit P.K. (2021). Spatio-temporal analysis of COVID-19 in India–a geostatistical approach. Spatial Information Research, 29, 661-672, DOI: 10.1007/s41324-020-00376-0.

Choi J.Y. (2020). COVID-19 in South Korea. Postgraduate medical journal, 96, 399-402, DOI: 10.1136/postgradmedj-2020-137738.

Cliff A.D., Ord J.K. (1981). Spatial processes: models & applications. Taylor & Francis.

Das A., Ghosh S., Das K., Basu T., Dutta I., Das M. (2021). Living environment matters: Unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. Sustainable Cities and Society 65, 102577, DOI: 10.1016/j.scs.2020.102577.

Duy C., Nong V.M., Van Ngo A., Doan T.T., Nguyen T.Q., Truong P.T. (2021). Nosocomial coronavirus disease outbreak containment, Hanoi, Vietnam, March–April 2020. Emerging infectious diseases, 27(1), 10-17, DOI: 10.3201/eid2701.202656.

Getis A., Ord J. K. (1996). Local spatial statistics: an overview. Spatial analysis: modelling in a GIS environment, 374, 261-277.

Ha B.T.T., La Quang N., Mirzoev T., Tai N.T., Thai P.Q., Dinh P.C. (2020). Combating the COVID-19 epidemic: experiences from Vietnam. International journal of environmental research and public health, 17, 3125, DOI: 10.3390/ijerph17093125.

Han Y., Yang L., Jia K., Li J., Feng S., Chen W. (2021). Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with environmental factors, Science of The Total Environment 761:144257, DOI: 10.1016/j.scitotenv.2020.144257.

HanoiTimes. (2021). Why does Hai Duong become a hotspot of Covid-19? http://hanoitimes.vn/why-does-hai-duong-become-a-hotspot-of-covid-19-316326.html. [Accessed 01 Nov. 2021].

Hoang H.A., Vu T.D., Nguyen T.T. (2017). Spatial Variability Analysis of Cu Content: A Case Study in Jiurui Copper Mining Area. International Journal of Applied Geospatial Research (IJAGR), 8,81-93, DOI: 10.4018/IJAGR.2017010105.

Hoang V.M., Hoang H.H., Khuong Q.L., La N.Q., Tran T.T.H. (2020). Describing the pattern of the COVID-19 epidemic in Vietnam. Global health action, 13, 1776526, DOI: 10.1080/16549716.2020.1776526.

Huang R., Xia J., Chen Y., Shan C., Wu C. (2020). Outbreak investigation for COVID-19 in northern Vietnam. The Lancet Infectious Diseases, 20(5), 535-536, DOI: 10.1016/S1473-3099(20)30159-6.

Kim S., Castro M.C. (2020). Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). International Journal of Infectious Diseases 98, 328-333, DOI: 10.1016/j.ijid.2020.07.004.

La V.P., Pham T.H., Ho M.T., Nguyen M.H., Nguyen K.L., Vuong T.T., Vuong Q.H. (2020). Policy response, social media and science journalism for the sustainability of the public health system amid the COVID-19 outbreak: the Vietnam lessons. Sustainability, 12(7), 2931, DOI: 10.3390/su12072931.

Lalor G.C., Zhang C. (2001). Multivariate outlier detection and remediation in geochemical databases Science of the total environment 281(1-3), 99-109, DOI: 10.1016/S0048-9697(01)00839-7.

Le T.H., Tran T.P.T. (2021). Alert for COVID-19 second wave: A lesson from Vietnam. Journal of global health, 11: 03012, DOI: 10.7189/jogh.11.03012.

Liu M. et al. (2021). The spatial clustering analysis of COVID-19 and its associated factors in mainland China at the prefecture level, Science of The Total Environment, 777, 145992, DOI: 10.1016/j.scitotenv.2021.145992.

MHV (2021). COVD-19 Information Website of Ministry of Health, Vietnam. Ministry of Health of Vietnam. https://ncov.moh.gov.vn/. [Accessed 01 Nov. 2021]

Mitchel A. (2005). The ESRI Guide to GIS analysis, Volume 2: Spartial measurements and statistics ESRI Guide to GIS analysis. ESRI press.

Nguyen T., Liu X., Ren Z. (2014). A Study Of Geochemical Exploration Spatial Cluster Identificaton Based On Local Spatial Autocorrelation. Geophysical and Geochemical Exploration. 38, 370-376 (In Chinese with English summary).

Nguyen T.H., Vu D.C. (2020). Summary of the COVID-19 outbreak in Vietnam–Lessons and suggestions. Travel medicine and infectious disease, 37, 101651, DOI: 10.1016/j.tmaid.2020.101651.

Nguyen T.T. (2018). Use of Moran's I and robust statistics to separate geochemical anomalies in Jiurui area (Southeast China). Bulletin Of The Mineral Research and Exploration, 156(156), 179-192, DOI: 10.19111/bmre.01695.

Nguyen T.T., Vu D.T., Trinh L.H., Nguyen T.L.H. (2016). Spatial Cluster and Outlier Identification of Geochemical Association of Elements: A Case Study in Juirui Copper Mining Area. Bulletin Of The Mineral Research and Exploration, 153(153), 159-167, DOI: 10.19111/bulletinofmre.351376.

Nguyen T.T., Vu T.D. (2019a). Identification of multivariate geochemical anomalies using spatial autocorrelation analysis and robust statistics. Ore Geology Reviews, 111:102985, DOI: 10.1016/j.oregeorev.2019.102985.

Nguyen T.T., Vu T.D. (2019b). 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. Environment and Natural Resources Journal, 17(3), 1-10, DOI: 10.32526/ennrj. 32517.32523. 32019.32517.

Nguyen T.V., Dai Tran Q., Phan L.T., Vu L.N., Truong D.T.T., Truong H.C., Pham Q.D. (2021). In the interest of public safety: rapid response to the COVID-19 epidemic in Vietnam. BMJ global health, 6(1), e004100, DOI: 10.1136/bmjgh-2020-004100.

Ord J.K., Getis A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. Geographical analysis, 27, 286-306, DOI: 10.1111/j.1538-4632.1995.tb00912.x.

Phuong H.V.M., Dai Tran Q., Phan L.T., Vu L.N., Truong D.T.T., Truong H.C., Pham Q.D. (2021). Novel Mutation of SARS-CoV-2, Vietnam, July 2020. Emerging Infectious Diseases, 27(1519), DOI: 10.1136/bmjgh-2020-004100.

Ramírez-Aldana R., Gomez-Verjan J. C., Bello-Chavolla O. Y. (2020). Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. PLoS neglected tropical diseases, 14:e0008875, DOI: 10.1371/journal.pntd.0008875.

Reuters (2021). Vietnam taps private hospitals as Delta-driven COVID-19 infections rise. Reuters. https://www.reuters.com/world/asia-pacific/vietnam-taps-private-hospitals-delta-driven-covid-19-infections-rise-2021-07-30/. [Accessed 01 Nov. 2021].

Robinson T.P. (2000). Spatial statistics and geographical information systems in epidemiology and public health. Advances in Parasitology, 47, 81-128, DOI: 10.1016/S0065-308X(00)47007-7

Rose-Redwood R. et al. (2020). Geographies of the COVID-19 pandemic Dialogues in Human. Geography, 10(2), 97-106.

Shepherd M. (2020). Why Geography is a key-part of fighting the COVID-19 Coronavirus outbreak. Forbes.

Tough R. (2021). Ho Chi Minh City during the fourth wave of COVID-19 in Vietnam. City & Society 33(3), 1-13. Available at: https://ueaeprints.uea.ac.uk/id/eprint/81878/1/Accepted_Manuscript.pdf [Accessed 01 Nov. 2021]

Tran A., Deparis X., Dussart P., Morvan J., Rabarison P., Remy F., Gardon J. (2004). Dengue spatial and temporal patterns, French Guiana, 2001. Emerging infectious diseases, 10(615), DOI: 10.3201/eid1004.030186.

United Nations (2020). The Social Impact of COVID-19. https://www.un.org/development/desa/dspd/2020/04/social-impact-of-covid-19/. [Accessed 01 Nov. 2021].

VGP (2021). Gov't may impose stricter social distancing in COVID-19 hotspots. http://news.chinhphu.vn/Home/Govt-may-impose-stricter-social-distancing-in-COVID19-hotspots/20217/44715.vgp. [Accessed 01 Nov. 2021].

VietNamNews (2021a). Bắc Giang sees highest daily cases amid COVID-19 fourth wave. https://vietnamnews.vn/society/956680/bac-giang-sees-highest-daily-cases-amid-covid-19-fourth-wave.html. [Accessed 01 Nov. 2021].

VietNamNews (2021b). Fourth wave of COVID-19 in Việt Nam longer and much more serious than previous ones: Health Minister. VietNamNews. https://vietnamnews.vn/society/994206/fourth-wave-of-covid-19-in-viet-nam-longer-and-much-more-serious-than-previous-ones-health-minister.html. [Accessed 01 Nov. 2021].

VnExpress (2021). HCMC detects four new Covid clusters with unidentified transmission sources. https://e.vnexpress.net/news/news/hcmc-detects-four-new-covid-clusters-with-unidentified-transmission-sources-4292516.html. [Accessed 01 Nov. 2021].

VnExpress (2021a). Van Don airport in Covid-19 cluster reopens after more than a month. https://e.vnexpress.net/news/news/van-don-airport-in-covid-19-cluster-reopens-after-more-than-a-month-4242771.html. [Accessed 01 Nov. 2021].

VnExpress (2021b). Vietnam records nine more Covid-19 community transmissions. https://e.vnexpress.net/news/news/vietnam-records-nine-more-covid-19-community-transmissions-4231748.html. [Accessed 01 Nov. 2021].

VnExpress. (2021c). Covid-19 hotspot Bac Ninh permits reopening of Canon factory. https://e.vnexpress.net/news/business/companies/covid-19-hotspot-bac-ninh-permits-reopening-of-canon-factory-4281449.html. [Accessed 01 Nov. 2021].

Wang H., Du Z., Wang X., Liu Y., Yuan Z., Liu Y., Xue F. (2015). Detecting the association between meteorological factors and hand, foot, and mouth disease using spatial panel data models. International Journal of Infectious Diseases 34:66-70, DOI: 10.1016/j.ijid.2015.03.007.

WHO (2021). WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/. [Accessed May 2021].

Xie Z., Qin Y., Li Y., Shen W., Zheng Z., Liu S. (2020). Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. Science of The Total Environment 744:140929. DOI: 10.1016/j.scitotenv.2020.140929.

Zhang S.L., Zhang K. (2007). Comparison between general Moran's Index and Getis-Ord general G of spatial autocorrelation. Acta Scientiarum Naturalium Universitatis Sunyatseni, 4(022).