# Identifying Spatio-Temporal Clustering of the COVID-19 Patterns Using Spatial Statistics:

## Case Studies of Four Waves in Vietnam

Anh-huy Hoang, Hanoi University of Natural Resources and Environment, Vietnam
Tien-thanh Nguyen, Hanoi University of Natural Resources and Environment, Vietnam*
https://orcid.org/0000-0002-7600-5858

### ABSTRACT

An outbreak of the COVID-19 pandemic caused by the SARS CoV 2 has profoundly affected the world. This study aimed to identify the spatio-temporal clustering of COVID-19 patterns using spatial statistics. Local Moran's I spatial statistic and Moran scatterplot were first used to identify high-high and low-low clusters and low-high and high-low outliers of COVID-19 cases. Getis-Ord's⟦ G⟧_i^* statistic was then applied to detect hotspots and coldspots. The authors finally illustrated the used method by using a dataset of 10,742 locally transmitted cases in four COVID-19 waves in 63 prefecture-level cities/provinces in Vietnam. The results showed that significant low-high spatial outliers of COVID-19 cases were first detected in the north-eastern region in the first wave and in the central region in the second wave whereas spatial clustering of high-high, low-high, and high-low was mainly found in the north-eastern region in the last two waves. It can be concluded that spatial statistics are of great help in understanding the spatial clustering of COVID-19 patterns.

### KEYWORDS

COVID-19 Pandemic, Spatial Patterns, Spatial Statistics, Spatio-Temporal Clustering, Vietnam

### INTRODUCTION

The COVID-19 pandemic, caused by the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been described as an unprecedented global health and socio-economic crisis (Kassema, 2020). The latest data show that, globally, as of 18 June 2021, the COVID-19 pandemic has resulted in more than 177.1 million confirmed cases and more than 3.8 million deaths (WHO, 2021). The pandemic will create socio-economic burdens differently in developed and developing counties of the World due to the loss of human resources (United Nations, 2020). It is, therefore, the use of modern techniques, especially geographical approaches, is of great help in the fight against the COVID-19 pandemic in general, and in the understanding of the spatial distribution and managing the COVID-19 infection in particular.

DOI: 10.4018/IJAGR.297517 *Corresponding Author

COVID-19-related data such as the geographical locations of (visited) COVID-19 cases which have a spatial and geographic dimension can be considered a type of spatial object and can be studied with the help of a Geographic Information System (GIS) and spatial statistics. The clustering phenomenon of these spatial objects in general and COVID-19 infection in particular, itself is tied to Tobler's First Law of Geography, which offers that while all objects are related in space - objects nearer to one another are more related (Sadler & Furr-Holden, 2019, Tobler, 2004). It is, therefore, geographical approaches are fundamental to keep infectious diseases and their geographical distribution under control (Cicalò & Valentino, 2019). Following the idea of Tobler (2004), commonly used statistics for spatial auto-correlation analysis such as global spatial statistics (*Moran's I*, *Getis-Ord G** and *Geary's c*) and local indicators of spatial association (LISA) have been successfully applied in epidemiological studies (see Robinson (2000) for a detailed review) in general and in the study of COVID-19 pandemic (see the following review) in particular. Among these spatial statistics, *Getis-Ord G** and *Moran's I* have been widely used in the study of the COVID-19 pandemic. The *Getis-Ord G** statistic has proven its effectiveness in hotspot detection in epidemiology studies, thus, it has been widely applied to detect the COVID-19 hotspots at global, regional, country, and province-levels.

At a global level, with the combination of the global, local *Moran's I* and *Getis-Ord G** statistics, Fatima et al. (2021) successfully detected the spatial clustering and hotspots of COVID incidence in 2020. Later, when performing spatio-temporal analysis and hotspots detection of COVID-19 using geographic information system (GIS) with new confirmed COVID-19 cases collected at the end of March and April 2020, Shariati et al. (2020) also revealed that hotspot analysis coupled with local *Moran's I* provide a scrupulous and objective approach to determine the locations of statistically significant clusters of COVID-19 cases.

At a country level, extensive studies on the identification of COVID-19 hotspots based on *Getis-Ord G** statistic have been conducted in many badly-affected countries by COVID-19 such as China (Liu et al., 2021, Wang et al., 2021), United States (Mollalo et al., 2020), Brazil (Alves et al., 2021), Italy (Ghosh & Cartone, 2020), England (Sartorius et al., 2021), and most recently India (Bhunia et al., 2021, Parvin et al., 2021). One of the first studies on the spatial clustering of the COVID-19 pandemic was conducted by Wang et al. (2021) utilizing the global and local spatial correlation analysis (*Moran's I*). In that study, Wang et al. (2021) successfully detected the spatio-temporal changes of COVID-19 transmission in China using 81,000 COVID-19 confirmed case data collected in 337 prefecture-level cities from January to March 2020 at the beginning of the epidemic. Later, when analyzing temporal and spatial COVID-19 transmission in China, *Moran's I*, *Getis-Ord G**, *Kulldorff's* space-time scan statistics, and the geographically weighted regression model have proven their effectiveness in a study by Liu et al. (2021) to identify spatial clusters of 67,449 COVID-19 laboratory-confirmed cases and to investigate the effects of the associated factors on COVID-19 incidence, respectively. In the United States, the combination of *Getis-Ord G** statistic and artificial neural network in a GIS environment was employed by Mollalo et al. (2020) to model COVID-19 incidence rates across the Continental United States. Sartorius et al. (2021) effectively applied a Bayesian hierarchical space-time SEIR model to assess the spatio-temporal variability of COVID-19 caseloads (transmission) and deaths at a small-area scale in England. When analyzing COVID-19 spread in Iran, Ramírez-Aldana et al. (2020) successfully employed univariate and bivariate spatial statistics to identify spatial clustering of COVID-19 cases at a province level. With the aims of analyzing the geographical distribution of COVID-19 and of identifying high-risk areas for the occurrence of cases and deaths from the disease in the indigenous population of Brazil (Alves et al., 2021), both of *Getis-Ord General G* and *Getis-Ord G**statistics were successfully used by Alves et al. (2021) to identify high spatial risk clusters from 32,041 cases of and 471 deaths by COVID-19. In Italy, widely used spatial indicators such as the local *Moran's I*, *Getis and*