

## Detection of Hotspots of Tuberculosis Cases in Vietnam

Thi-Tuyet-Mai Nguyen<sup>1\*</sup>, Thi-Bich-Thuy Luong<sup>2</sup>

<sup>1</sup> Faculty of Pharmacy, East Asia University of Technology, Hanoi, Vietnam

<sup>2</sup> Faculty of Nursing, East Asia University of Technology, Hanoi, Vietnam

---

### ABSTRACT:

**Background:** Tuberculosis (TB) is regarded as one of the leading causes of death globally. It remains a significant cause of morbidity and mortality in Vietnam. This study aims to identify hotspots of TB using boxplot and Getis-Ord's  $G_i^*$  statistic-based hotspot analysis.

**Data used and Methods:** A total of 101,438 TB cases in 2020 collected from 63 provinces/cities in Vietnam was used in study. Boxplot is first used to study distribution of TB cases. Getis-Ord's  $G_i^*$  statistic was then employed to identify hotspots of TB cases. Finally, results and main findings will be discussed and concluded.

**Results:** It was found that a total of 05 hotspots and 04 coldspots of TB cases were detected throughout Vietnam. Five hotspots were detected in 05 provinces in the northeastern region including Ha Nam, Nam Dinh, Hai Phong, Hai Duong, and Hung Yen. Whereas, four coldspots were mainly concentrated in 03 provinces in the northwest region (Cao Bang, Tuyen Quang and Son La), and Dak Lak in the central south region.

**Conclusion:** It can be concluded that the combination of boxplot and Getis-Ord's  $G_i^*$  statistic can help to effectively detect hotspots of TB cases. Findings in this study provide an insight into how to use spatial statistics and spatial analysis in the study of TB distribution.

**KEYWORDS:** Detection, Hotspot analysis, Spatial autocorrelation, Tuberculosis, Local Getis Ord's statistic, Vietnam.

---

### INTRODUCTION

Tuberculosis (TB), an infectious disease caused by *Mycobacterium tuberculosis*, continues to infect many people in spite of global attempts to control the disease (1). The infectious illness tuberculosis (TB) is airborne and spreads from person to person through the respiratory system when a new host inhales aerosolized TB particles from an infected patient's cough (2). Among the top 10 causes of death worldwide, TB is ranked higher than HIV/AIDS, with 8 million cases of active TB reported annually (3). Its origins can be traced back to a class of fast-growing bacteria known as *Mycobacterium tuberculosis* complex. *M. tuberculosis sensu stricto* (MTBss) and *M. africanum* (MAF), which further separated into seven lineages (L), are the principal causes of tuberculosis in humans. MTBss subdivided into L1–L4 and L7, while MAF L5 and L6 (2, 3) (4). In 1993, the World Health Organization (WHO) designated tuberculosis (TB) as a worldwide health emergency, necessitating further funding and investigation to successfully manage the illness. Because of its low rate of recovery and co-infection with other illnesses, tuberculosis remains one of the largest threats to public health (4). WHO expected that about 10 million cases of TB were diagnosed and about 1.41 million deaths in 2019 globally. Southeast Asia makes up about 44% among other diseases (5). According to the WHO, 10 million people suffer from TB across the world in 2018, incurring 1.5 million deaths (6). To reach the ultimate goal of eliminating TB by 2050 (defined as  $\leq 1$  case per 1 million persons) (7). It is therefore, the identification of hotspots of TB plays a very important role in the fight of TB disease. At the global, national, local, and district levels, there are variations in the geographic distribution of tuberculosis. Indicating an unequal distribution of the disease burden across spatial scales is the geographic heterogeneity of tuberculosis (8). Policymakers may find it useful to have an understanding of the spatial distribution of tuberculosis cases and high-risk locations. This understanding can help inform public health actions, such as the customization and targeting of interventions (9). Finer spatial resolution of TB data is required in order to create successful interventions for local zones; this means moving away from solely examining nationally aggregated data and toward using pertinent local data (10). The better utilization of already-existing data, such as program reports of tuberculosis cases, and the extension of that routine data to incorporate spatial information are crucial to this change (9). Previous studies have demonstrated how evaluation of spatial variations of TB disease can identify high-risk areas to be spatially targeted with screening interventions (1). Spatial patterns of tuberculosis prevalence in many nations with diverse

## Detection of Hotspots of Tuberculosis Cases in Vietnam

environmental settings have been studied by the application of GIS and spatial statistics (11–16). These investigations have produced valuable data regarding the disease's distribution and modes of transmission. The primary drawback of these research, meanwhile, is their inclination toward smaller spatial scales, like a single urban region across brief temporal periods (17,18). These studies miss the geographical pattern of tuberculosis (TB) at a spatial scale that matters for TB programming, like the district size, because they concentrate on lower spatial scales. In Zimbabwe, the district serves as the geographical epidemiological administrative unit where plans for tuberculosis interventions and control are made (19). When the spatial unit at which TB data is analyzed reflects the anticipated epidemiological dynamics in the nation, spatial analysis becomes pertinent in this regard. Therefore, in order to guide targeted public health response—which is considered to be one of the most effective techniques in disease control - it is crucial to thoroughly understand the spatial variability in the prevalence of tuberculosis at large spatial scales, such as the district level (20). TB remains a major public health problem in developing countries such as Vietnam. Although, Vietnam has made significant progress in its TB control efforts, achieving Millennium Developmental Goal (MDG) targets for reductions in TB incidence, prevalence and mortality well before the 2015 target date (5). However, rising numbers of MDR-TB patients is a major concern; Vietnam is ranked 14th among the 27 highest MDR-TB burden countries in the world (5). Therefore, this study aims to identify hotspots of TB using boxplot and Getis-Ord's  $G_i^*$  statistic-based hotspot analysis. Boxplot is first used to study distribution of TB cases. Getis-Ord's  $G_i^*$  statistic was then employed to identify hotspots of TB cases. Finally, study results and main findings will be discussed and concluded.

### DATA USED

Vietnam has a low prevalence of multidrug-resistant TB (2.3%) in 1996-1997 (21). It is the only country of 22 countries with the highest number of TB cases worldwide that has reached and exceeded WHO targets for TB control since 1997 (22,23). In 2014, Vietnam was one of the high TB burden countries within the Asia-Pacific region, ranking 11th among the 22 highest TB burden countries in the world (5). TB is the second most common infectious cause of death in Vietnam, after respiratory infections in general, with 100,349 new and recurrent TB cases and 17,000 TB deaths recorded in 2014 (5). In this study, a total of 101,438 TB cases in Vietnam in 2020 will be used to detect hotspots of TB using boxplot and local Getis-Ord's  $G_i^*$  statistic-based hotspot analysis in this study. These TB cases were collected from websites of the Vietnam Ministry of Health (VMH).

### METHODS

#### Boxplot

In descriptive statistics, a box plot or boxplot is a method for graphically demonstrating the locality, spread and skewness groups of numerical data through their quartiles (24). In addition to the box on a box plot, there can be lines (which are called whiskers) extending from the box indicating variability outside the upper and lower quartiles, thus, the plot is also called the box-and-whisker plot and the box-and-whisker diagram. Outliers as shown in Figure 1 that differ significantly from the rest of the dataset may be plotted as individual points beyond the whiskers on the box-plot (25). Box plots are non-parametric: they display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution (though Tukey's boxplot assumes symmetry for the whiskers and normality for their length) (26). The spacings in each subsection of the box-plot indicate the degree of dispersion (spread) and skewness of the data, which are usually described using the five-number summary. In addition, the box-plot allows one to visually estimate various L-estimators, notably the interquartile range, midhinge, range, mid-range, and trimean. Box plots can be drawn either horizontally or vertically.

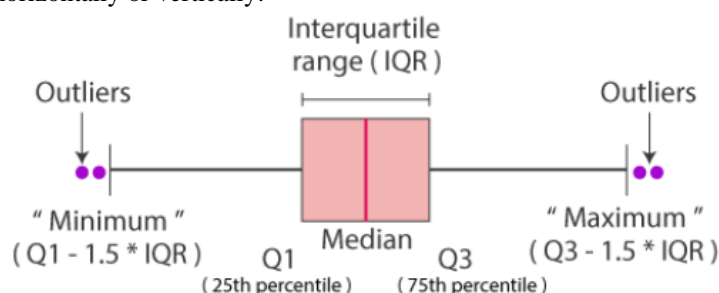


Figure 1. Different parts of boxplot.

A boxplot is a standardized way of displaying the dataset based on the five-number summary: the minimum, the maximum, the sample median, and the first and third quartiles. Minimum ( $Q_0$  or 0<sup>th</sup> percentile): the lowest data point in the data set excluding any outliers. Maximum ( $Q_4$  or 100<sup>th</sup> percentile): the highest data point in the data set excluding any outliers. Median ( $Q_2$  or 50<sup>th</sup> percentile): the middle value in the data set. First quartile ( $Q_1$  or 25<sup>th</sup> percentile) also known as the lower quartile  $q_n(0.25)$ , it is the median of the lower half of the dataset. Third quartile ( $Q_3$  or 75<sup>th</sup> percentile) also known as the upper quartile  $q_n(0.75)$ , it is the median of the upper half of the dataset (27). In addition to the minimum and maximum values used to construct a box-plot, another important element that can also be employed to obtain a box-plot is the interquartile range (IQR).

## Detection of Hotspots of Tuberculosis Cases in Vietnam

$$IQR = Q_3 - Q_1 = q_n(0.75) - q_n(0.25) \quad (1)$$

### Getis Ord $G_i^*$ statistic-based hotspot analysis

A hotspot can be defined as an area that has higher concentration of events compared to the expected number given a random distribution of events. Hotspot analysis is a spatial analysis and mapping technique interested in the identification of clustering of spatial phenomena. Hotspot analysis is based on the Getis-Ord's  $G_i^*$  statistic. 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 (28). Hotspot can separate clusters of high values from cluster of low values. It is, therefore, Getis-Ord's  $G_i^*$  statistic was used to identify the counties of high and low numbers of COVID-19 cases (29,30). The form of Getis-Ord's  $G_i^*$  statistic is defined as follows (31):

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

with:

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad (3)$$

and:

$$S = \sqrt{\frac{\sum_{j=1}^N x_j^2}{N} - (\bar{x})^2} \quad (4)$$

Expectation:

$$E(G_i^*) = \frac{W_i^*}{n-1} \quad (5)$$

with:

$$W_i^* = \sum_{j=1}^n w_{ij}(d) \quad (6)$$

and variance:

$$\text{Var}(G_i^*) = \frac{W_i^*(n - W_i^*)Y_{i2}^*}{n^2(n-1)(Y_{i1}^*)^2} \quad (7)$$

with:

$$Y_{i1}^* = \frac{\sum_{j=1}^n x_j}{n}; \quad (8)$$

and:

$$Y_{i2}^* = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_i x_j)^2}{n} - (Y_{i1}^*)^2; \quad (9)$$

where: the Getis-Ord's  $G_i^*$  statistic is computed for the number of COVID-19 cases at county  $i$ ;  $x_i$ ,  $x_j$ ,  $\bar{x}$ , and  $W_{ij}$  are spatial weight matrix; 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.

The Getis-Ord  $G_i^*$  statistic returned for each feature in the dataset is a Z-score. For statistically significant positive Z-scores, the larger the Z-score is, the more intense the clustering of high values (hotspot). For statistically significant negative Z-scores, the smaller the Z-score is, the more intense the clustering of low values (coldspot). The application of hotspot analysis within public health and epidemiological research as well as in other disciplines (e.g., a great deal of the literature on hotspot analysis comes from crime mapping and research) has increased significantly in the past couple of decades mainly due to the advent of geographic information systems (GIS)-based software.

## RESULTS AND DISCUSSION:

### Spatial distribution of Tuberculosis cases

The data from the boxplot and map in Figure 2 show the distribution of TB cases in Vietnam in 2020. The data from the boxplot in Figure 2-right shows the smallest and largest number of cases were 117 cases and 17,676 cases, respectively. The values of the first (1-st quartile -  $Q_1$ ) and third (3-rd quartile -  $Q_3$ ) interquartile ranges were 647 and 1722, respectively. Meanwhile, the mean and median values of TB cases were 1,530 and 1,109 cases, respectively. The dispersion of the variable of number of infections with

## Detection of Hotspots of Tuberculosis Cases in Vietnam

IQR and SD parameters was 1075.5 and 2264.7, respectively. It can be seen that the mean value was larger than the median. Whereas the SD dispersion value was larger than the IQR. It is therefore, the boxplot tends to skew toward the top as shown in Figure 2-left.

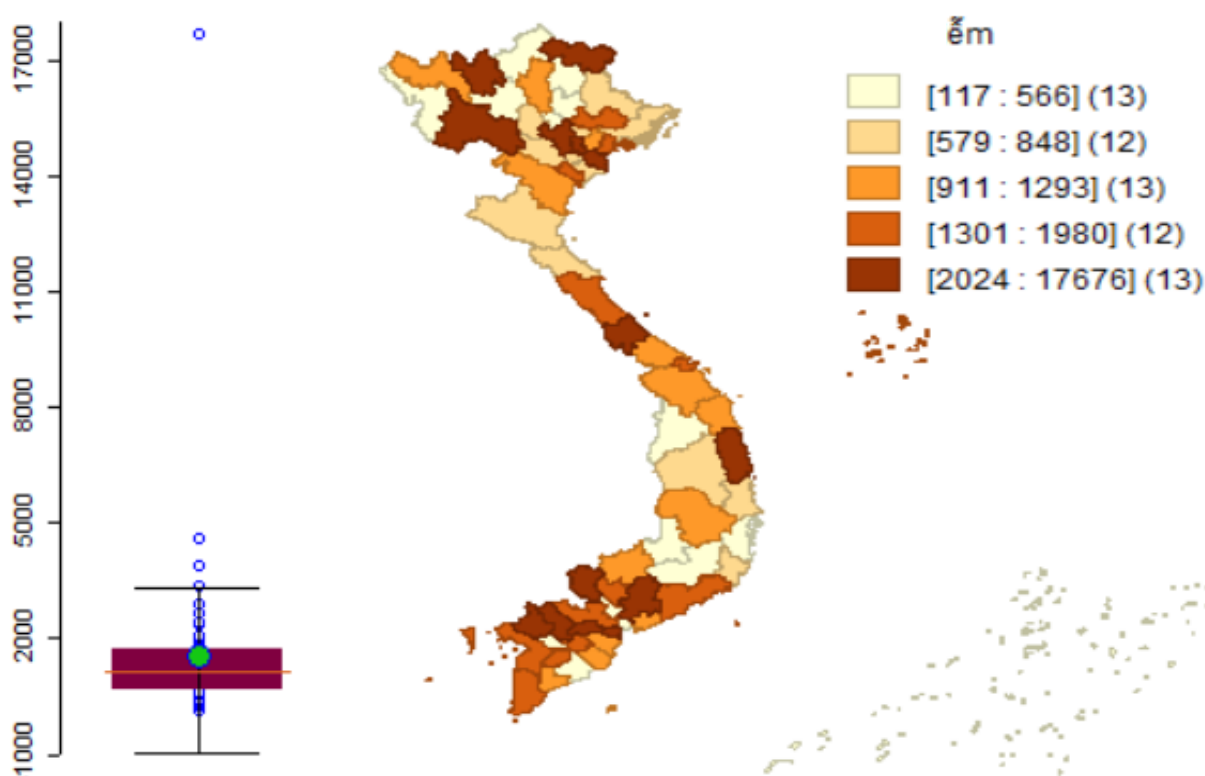


Figure 2. Boxplot (left) and map (right) of TB cases in Vietnam.

Data from the map in Figure 2-right shows the spatial distribution of TB cases throughout Vietnam. The number of TB cases is classified into five ranges: very low (from 117 to 566 cases), low (from 579 to 848 cases), medium (from 911 to 1293 cases), high (from 1301 to 1980 cases) and very high (from 2024 to 17676 cases). It can be seen from the map of distribution of TB cases in 2020 in Figure 2-right, a large number of TB cases were mainly concentrated in the northern region, the central region and some provinces in the southwest region of Vietnam. A large number of TB cases was detected in some provinces/cities such as Ho Chi Minh City (17,676 cases), followed by An Giang (4,611 cases), Hanoi (3,910 cases), Dong Nai (3,345 cases), Binh Dinh (2892 cases), Hung Yen (2693 cases). The province with the lowest number of TB cases is Kon Tum (117 cases), followed by Bac Kan (134 cases), Vinh Phuc (152 cases), Dien Bien (161 cases), and Dak Nong (278 cases).

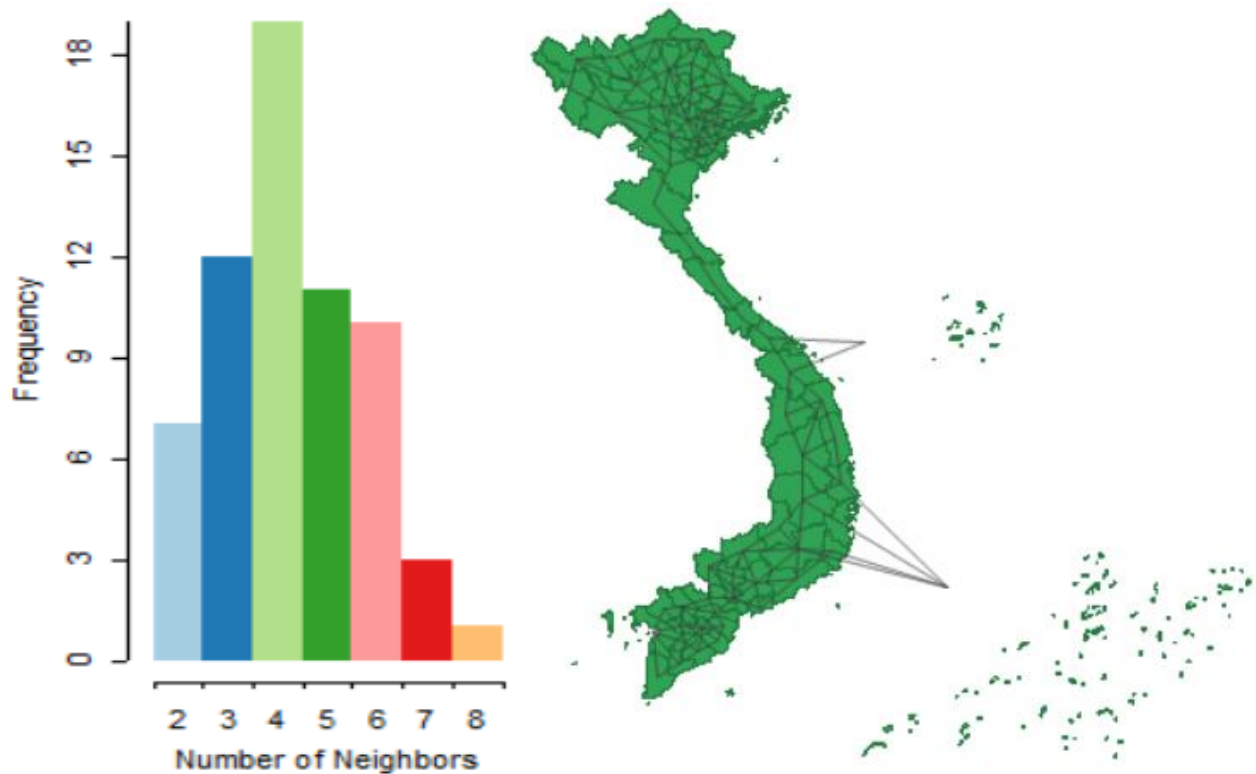
### Analysis of spatial distribution of Getis-Ord's statistic

Data from Table 1 summarizes the attribute values of the spatial weight matrix used to compute the Getis-Ord's  $G_i^*$  statistic. It can be seen from data from Table 1 that the total observations (all provinces/cities) was 63. The province having the least and maximum number of neighbors was 2 and 8, respectively. Meanwhile, throughout the territory of Vietnam, the mean and median values for the number of neighbors were 4.29 and 4, respectively. The distribution of neighbors was shown in Figure 3-left. Data in Figure 3-left shows that the number of neighbors of provinces/cities throughout the territory of Vietnam was distributed quite evenly. The data was evenly distributed on both sides, in which, a total of 18 provinces/cities had 4 neighbors. The network of neighbors of provinces/cities was shown in Figure 3-right.

Table 1. Summary of spatial weight matrix.

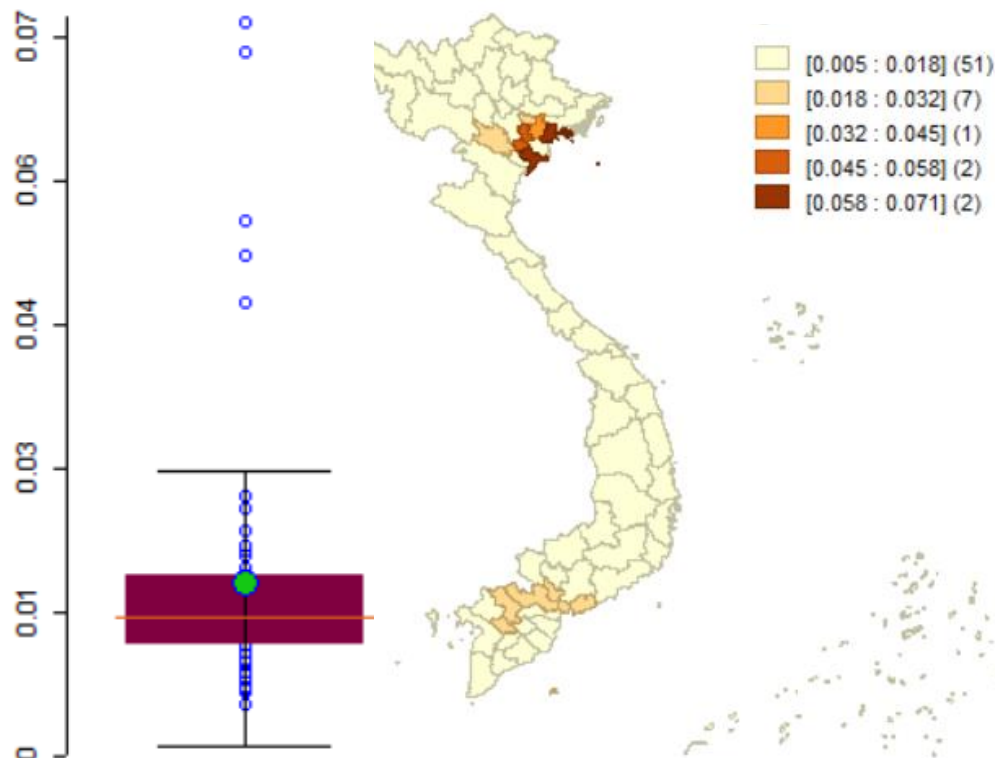
| Properties       | Values |
|------------------|--------|
| Observations     | 63     |
| Min neighbors    | 2      |
| Max neighbors    | 8      |
| Mean neighbors   | 4.29   |
| Median neighbors | 4      |
| %Non-zero        | 6.8%   |

## Detection of Hotspots of Tuberculosis Cases in Vietnam



**Figure 3.** Histogram of neighbors of and connectivity map used to construct spatial weight matrix.

The data from the boxplot and map in Figure 4 shows the distribution of the Getis-Ord's  $G_i^*$  statistic of TB cases in Vietnam in 2020. The data from the boxplot in Figure 4-left shows that the smallest and largest values of the Getis-Ord's  $G_i^*$  statistic were 117 cases and 17,676 cases, respectively. The values of the first and third quartiles was 0.011 and 0.018, respectively. Meanwhile, The mean and median values of Getis-Ord's  $G_i^*$  statistic were 0.017 and 0.013 respectively. The dispersion values of Getis-Ord's  $G_i^*$  statistic, interquartile ranges (IQR) and standard deviation (SD) parameters, were 0.007 and 0.013, respectively. It can be seen that the mean value was equivalent to the median value. Meanwhile, the SD dispersion value was larger than the IQR, which makes the boxplot also tend to skew toward the top as shown in Figure 4-left.



**Figure 4.** Boxplot (left) and map (right) of Getis-Ord's  $G_i^*$  statistic obtained from TB cases in Vietnam.

## Detection of Hotspots of Tuberculosis Cases in Vietnam

Data from the Map in Figure 4-right shows the spatial distribution of the Getis-Ord's  $G_i^*$  statistic determined from the number of TB cases throughout Vietnam in 2020. Values of Getis-Ord's  $G_i^*$  statistic were divided into five different ranges: very low (from 0.05 to 0.018), low (from 0.018 to 0.032), medium (from 0.032 to 0.045), high (from 0.045 to 0.058). It can be seen from the Getis-Ord's  $G_i^*$  statistic distribution map in Figure 3-right that high values of Getis-Ord's  $G_i^*$  statistic were mainly concentrated in provinces/cities in the Northeast and Southwest regions of the country of Vietnam. It can be seen from the map in Figure 3-right that high values of the Getis-Ord's  $G_i^*$  statistic was mainly concentrated in some provinces/cities having a large number of TB cases in 2020 such as Ho Chi Minh City (17,676 cases), and followed by An Giang (4611 cases), Hanoi (3910 cases), Dong Nai (3345 cases), Binh Dinh (2892 cases), Hung Yen (2693 cases). Whereas, low values of the Getis-Ord's  $G_i^*$  statistic were mainly detected in the northwest, the north of the north, the central region and the southeast. These low values of Getis-Ord's  $G_i^*$  statistic were concentrated in provinces with a low number of TB infections such as Kon Tum (117 cases), then Bac Kan (134 cases), Vinh Phuc (152 cases), Dien Bien (161 cases), and Dak Nong (278 cases).

### Analysis of hotspots of Tuberculosis cases

The results of identifying hotspots from the number of TB cases in Vietnam in 2020 are shown in Figure 5. Data from the Getis-Ord's  $G_i^*$  statistic cluster map in Figure 5-left shows that there were a total of 05 hotspots and 04 coldspots of TB cases were detected throughout Vietnam. Specifically as follows, 05 hotspots were discovered in 05 provinces in the northeastern region including Ha Nam (848 cases), Nam Dinh (785 cases), Hai Phong city (1578 cases), Hai Duong (1202 cases), Hung Yen (2693 cases). Whereas, 04 coldspots were concentrated in 03 provinces in the northwest region and 01 province in the central south region. Specifically, they included Cao Bang (2,350 cases), Tuyen Quang (1,125 cases), Son La (2,430 cases), and Dak Lak (1,109 cases). The remaining 54 provinces/cities were found with no hotspots or coldspots for the number of TB cases. These provinces/cities are mainly concentrated in the central and southeastern regions. Some provinces/cities had a relatively high number of TB cases but no hotspots in the number of TB cases were detected such as Ho Chi Minh City (17,676 cases), An Giang (4,611 cases), Ha Noi (3910 cases), Dong Nai (3345 cases), Binh Dinh (2892 cases). Similarly, some provinces having a low number of TB cases but no coldspots were detected such as Kon Tum (117 cases), Bac Kan (134 cases), Vinh Phuc (152 cases), Dien Bien (161 cases), and Dak Nong (278 cases).

The data from Figure 5-right shows the spatial distribution of the level of statistical significance (p-value) achieved by the local Getis-Ord's  $G_i^*$  statistic for each province/city in Vietnam. Statistical significance levels are expressed on four different scales: statistical insignificance (p-value > 0.05) and statistical significance at the levels of 0.05, 0.01 and 0.001. Data from Figure 5-right shows that local Getis-Ord's  $G_i^*$  statistic statistically significant at the 0.05 level was detected in 6 provinces/cities, including Nam Dinh (785 cases), Hai Phong city (1578 cases), Hai Duong (1202 cases), Hung Yen (2693 cases), Cao Bang (2350 cases), Son La (2430 cases). Whereas, the local Getis-Ord's  $G_i^*$  statistic with high significance at the 0.01 level was found in three provinces/cities, including Tuyen Quang (1125 cases), Ha Nam (848 cases), and Dak Lak (1109 cases).

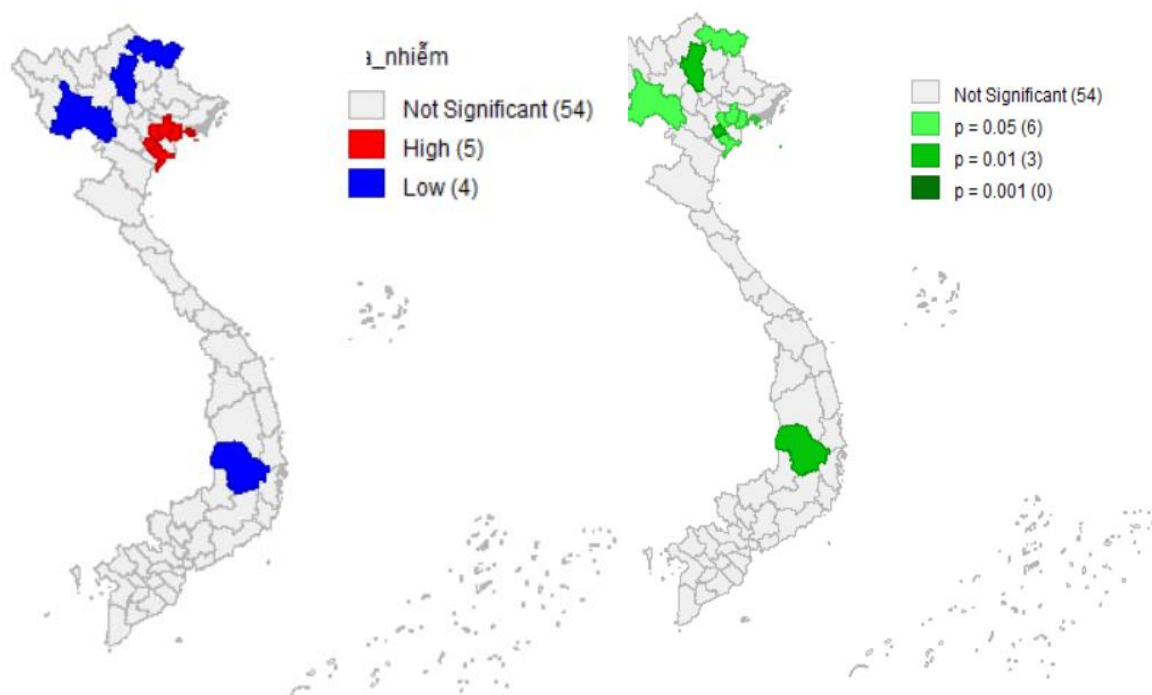


Figure 5. Hotspot (left) and significant (right) map of Tuberculosis cases in Vietnam.

## Detection of Hotspots of Tuberculosis Cases in Vietnam

### CONCLUSION

This study was conducted to identify hotspots of TB using boxplot and Getis-Ord's  $G_i^*$  statistic-based hotspot analysis. A total of 101,438 TB cases in 2020 collected from 63 provinces/cities in Vietnam was used in study. Boxplot was first used to study distribution of TB cases. Getis-Ord's  $G_i^*$  statistic-based hotspot analysis was then employed to identify hotspots of TB cases. Finally, results and main findings will be discussed and concluded. It was found that a total of 05 hotspots and 04 coldspots of TB cases were detected throughout Vietnam. Five hotspots were detected in 05 provinces in the northeastern region including Ha Nam, Nam Dinh, Hai Phong, Hai Duong, and Hung Yen. Whereas, four coldspots were mainly concentrated in 03 provinces in the northwest region (Cao Bang, Tuyen Quang and Son La), and Dak Lak in the central south region. It can be concluded that the combination of boxplot and Getis-Ord's  $G_i^*$  statistic can help to effectively detect hotspots of TB cases. Findings in this study provide an insight into how to use spatial statistics and spatial analysis in the study of TB distribution. Spatial statistics were successfully applied to determine the extent to which TB cases were clustered in space at province/city level in Vietnam. The information generated from this study is useful in providing detailed knowledge on the spatial pattern of TB occurrence which is critical for targeted TB interventions.

**ACKNOWLEDGEMENT:** The authors would like to thank the Vietnam Ministry of Health for providing data. We are extremely grateful to the editor and the anonymous reviewer for their valuable comments and suggestions, which have helped improve the quality of our manuscript.

**CONFLICT OF INTEREST:** None

### REFERENCES

- 1) Kiani B, Raouf Rahmati A, Bergquist R, Hashtarkhani S, Firouraghi N, Bagheri N, et al. Spatio-temporal epidemiology of the tuberculosis incidence rate in Iran 2008 to 2018. *BMC Public Health*. 2021;21:1–20.
- 2) Sia IG, Wieland ML. Current concepts in the management of tuberculosis. In: *Mayo Clinic Proceedings*. Elsevier; 2011. p. 348–61.
- 3) Alene KA, Viney K, Moore HC, Wagaw M, Clements ACA. Spatial patterns of tuberculosis and HIV co-infection in Ethiopia. *PLoS One*. 2019;14(12):e0226127.
- 4) Mohidem NA, Osman M, Hashim Z, Muharam FM, Mohd Elias S, Ma'pol A. Mapping hotspots of tuberculosis cases with validation on site in Gombak, Selangor, Malaysia. *Spat Inf Res*. 2023;1–12.
- 5) Organization WH. *Global tuberculosis report 2013*. World Health Organization; 2013.
- 6) MacNeil A, Glaziou P, Sismanidis C, Maloney S, Floyd K. Global epidemiology of tuberculosis and progress toward achieving global targets—2017. *Morb Mortal Wkly Rep*. 2019;68(11):263.
- 7) Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annu Rev Public Health*. 2013;34.
- 8) Cohen T, Colijn C, Finklea B, Murray M. Exogenous re-infection and the dynamics of tuberculosis epidemics: local effects in a network model of transmission. *J R Soc Interface*. 2007;4(14):523–31.
- 9) Theron G, Jenkins HE, Cobelens F, Abubakar I, Khan AJ, Cohen T, et al. Data for action: collection and use of local data to end tuberculosis. *Lancet*. 2015;386(10010):2324–33.
- 10) Brooks MB, Millones AK, Puma D, Contreras C, Jimenez J, Tzelios C, et al. Mapping local hot spots with routine tuberculosis data: A pragmatic approach to identify spatial variability. *PLoS One*. 2022;17(3):e0265826.
- 11) Tiwari N, Adhikari CMS, Tewari A, Kandpal V. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. *Int J Health Geogr*. 2006;5:1–11.
- 12) Touray K, Adetifa IM, Jallow A, Rigby J, Jeffries D, Cheung YB, et al. Spatial analysis of tuberculosis in an urban west African setting: is there evidence of clustering? *Trop Med Int Heal*. 2010;15(6):664–72.
- 13) Liu Y, Li X, Wang W, Li Z, Hou M, He Y, et al. Investigation of space-time clusters and geospatial hot spots for the occurrence of tuberculosis in Beijing. *Int J Tuberc Lung Dis*. 2012;16(4):486–91.
- 14) Onozuka D, Hagihara A. Geographic prediction of tuberculosis clusters in Fukuoka, Japan, using the space-time scan statistic. *BMC Infect Dis*. 2007;7:1–9.
- 15) Wang T, Xue F, Chen Y, Ma Y, Liu Y. The spatial epidemiology of tuberculosis in Linyi City, China, 2005–2010. *BMC Public Health*. 2012;12:1–8.
- 16) Randremanana RV, Sabatier P, Rakotomanana F, Randriamanantena A, Richard V. Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City. *Trop Med Int Heal*. 2009;14(4):429–37.
- 17) Corbett EL, Charalambous S, Moloji VM, Fielding K, Grant AD, Dye C, et al. Human immunodeficiency virus and the prevalence of undiagnosed tuberculosis in African gold miners. *Am J Respir Crit Care Med*. 2004;170(6):673–9.
- 18) Chirenda J, Gwitira I, Warren RM, Sampson SL, Murwira A, Masimirembwa C, et al. Spatial distribution of Mycobacterium tuberculosis in metropolitan Harare, Zimbabwe. *PLoS One*. 2020;15(4):e0231637.

## Detection of Hotspots of Tuberculosis Cases in Vietnam

- 19) Gwitira I, Murwira A, Mberikunashe J, Masocha M. Spatial overlaps in the distribution of HIV/AIDS and malaria in Zimbabwe. *BMC Infect Dis.* 2018;18:1–10.
- 20) Keshavjee S, Dowdy D, Swaminathan S. Stopping the body count: a comprehensive approach to move towards zero tuberculosis deaths. *Lancet.* 2015;386(10010):e46–7.
- 21) Organization WH. Anti-tuberculosis drug resistance in the world: third global report. World Health Organization; 2004.
- 22) Organization WH. Global tuberculosis control: surveillance, planning, financing: WHO report 2008. Vol. 393. World Health Organization; 2008.
- 23) Huong NT, Duong BD, Co N V, Quy HT, Tung LB, Bosman M, et al. Establishment and development of the national tuberculosis control programme in Vietnam. *Int J Tuberc Lung Dis.* 2005;9(2):151–6.
- 24) DuToit SHC, Steyn AGW, Stumpf RH. Graphical exploratory data analysis. Springer Science & Business Media; 2012.
- 25) Grubbs FE. Procedures for detecting outlying observations in samples. *Technometrics.* 1969;1–21.
- 26) Boddy R, Smith G. Statistical methods in practice: for scientists and technologists. John Wiley & Sons; 2009.
- 27) Holmes A. Introductory business statistics. OpenStax College; 2017.
- 28) 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).
- 29) 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.
- 30) 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).
- 31) Cliff AD, Ord JK. Spatial processes: models & applications. (No Title). 1981;