MORAN'S I SPATIAL AUTOCORRELATION OF COVID-19 INFECTION RATES: IS THERE SUCH IN BUTUAN CITY?

Cheryl T. Gallor[1], Jerome T. Remetar[1], and Kendel P. Bolanio[1, 2]
[1]Department of Geodetic Engineering, Caraga State University, Ampayon, Butuan City 8600, Philippines,
Email: cheryltangallor@gmail.com, jerome.remetar@carsu.edu.ph
[2]Caraga Center for Geo-informatics, Caraga State University, Ampayon, Butuan City 8600, Philippines,
Email: kendel.bolanio@gmail.com

KEY WORDS: COVID-19, Global Spatial Autocorrelation, Incremental Spatial Autocorrelation, Multi-Distance Clustering, Local Spatial Autocorrelation

ABSTRACT: Understanding the spatial autocorrelation of COVID-19 can aid in scientific decision-making in suppressing this infectious disease. This study calculated the cumulative incidence rate of barangays in Butuan City affected by the outbreak of the COVID-19 pandemic from April 15, 2020 to February 28, 2021. This research used Global statistics, i.e., Global Moran's I and Getis-Ord General G, to investigate if spatial autocorrelation occurs in the study region. This study also examines clustering by distance using incremental spatial autocorrelation and multi-distance spatial cluster analysis. The analysis of the hot spot/cold spot and Anselin Local Moran's I were then used to locate high- and low-risk areas of the COVID-19 pandemic. Barangays with a 90 – 95% confidence interval were identified as significant. Moreover, cluster and outlier analysis demonstrated that ten Barangays exhibit HH-cluster and appear to be high-risk clusters.

1. INTRODUCTION

COVID-19, also known as coronavirus 2019, rapidly increased as it was first reported in Wuhan, China, in early December 2019 and spread quickly, bringing the total cumulative cases of 22,112 as of February 2021 (Kang et al., 2020). On January 13, 2020, the first case of COVID-19 outside China was first recorded in Thailand, and the number of countries affected outside China is rapidly increasing. Since it affects 219 countries and territories worldwide, the World Health Organization (WHO) declares coronavirus disease 2019 pandemic (Ramirez-Aldana et al., 2020).

The Philippines is part of this global pandemic, which reported its first confirmed positive case of COVID-19 on January 30, 2020, and its first local transmission on March 7, 2020. As of November 12, 2020, the total cumulative confirmed COVID-19 case in the Philippines had reached up to 401,416, recovered cases of 362,217, and total deaths of 7,710 updated report by the Department of Health (DOH) COVID-19 Tracker website. COVID-19 reached in Caraga region on April 6, 2020, and reported its first confirmed case from Butuan City. Last June 19, 2020, government authorities declared a local transmission in the city. Because of this local transmission, Butuan City is now one of the Philippines' cities with high counts of COVID-19 cases and still increasing. This viral disease spread rapidly within the city.

The 2019 Coronavirus Disease (COVID-19) outbreak has had a massive effect on people's lives and socio-economic growth. Understanding the spatial trends and influencing factors of the multi-scale COVID-19 epidemic could help manage the outbreak (Xiong et al., 2020). Most of the studies in analyzing geographical and temporal patterns of the COVID-19 use the spatial autocorrelation method. This spatial-statistical approach enables us to understand and analyze the COVID-19 disease patterns to predict and determine the factors that affect the rapid spread of coronavirus 2019 (Kang et al., 2020, Ramirez-Aldana et al., 2020, Xiong et al., 2020, Shariati et al., 2020). The primary goal of this research is to determine the spatial association of COVID-19 infection rate in Butuan city with the use of a Geographic Information System and spatial statistics. Specifically, this study aims: (1) to generate maps and statistical graphs showing information of cumulative and confirmed COVID-19 cases, (2) to identify the spatial clusters of COVID-19 incidence using Moran's I spatial autocorrelation, and (3) to determine hot/cold spots and cluster/outlier of disease rates in the area using local spatial autocorrelation approaches.

2. METHODOLOGY

2.1 Overview of the Study

This study had investigated the spatial association of COVID-19 incidence rate in Butuan City using Moran's I. Global Moran's I was used to determining the spatial autocorrelation in the study region. Getis-Ord General G, Incremental Spatial Autocorrelation, and Ripleys-K function assessed Global Moran's I results. Local Getis-Ord Gi* and Local Moran's I identified spatial autocorrelation in a barangay scale.

2.2 Study Area

The study area is in Butuan City, shown in Figure 1. Butuan City is a highly urbanized city in the Caraga region, with a land area of 816.62 square kilometers or 315.30 square miles and a population of 337,063 based on Census data.

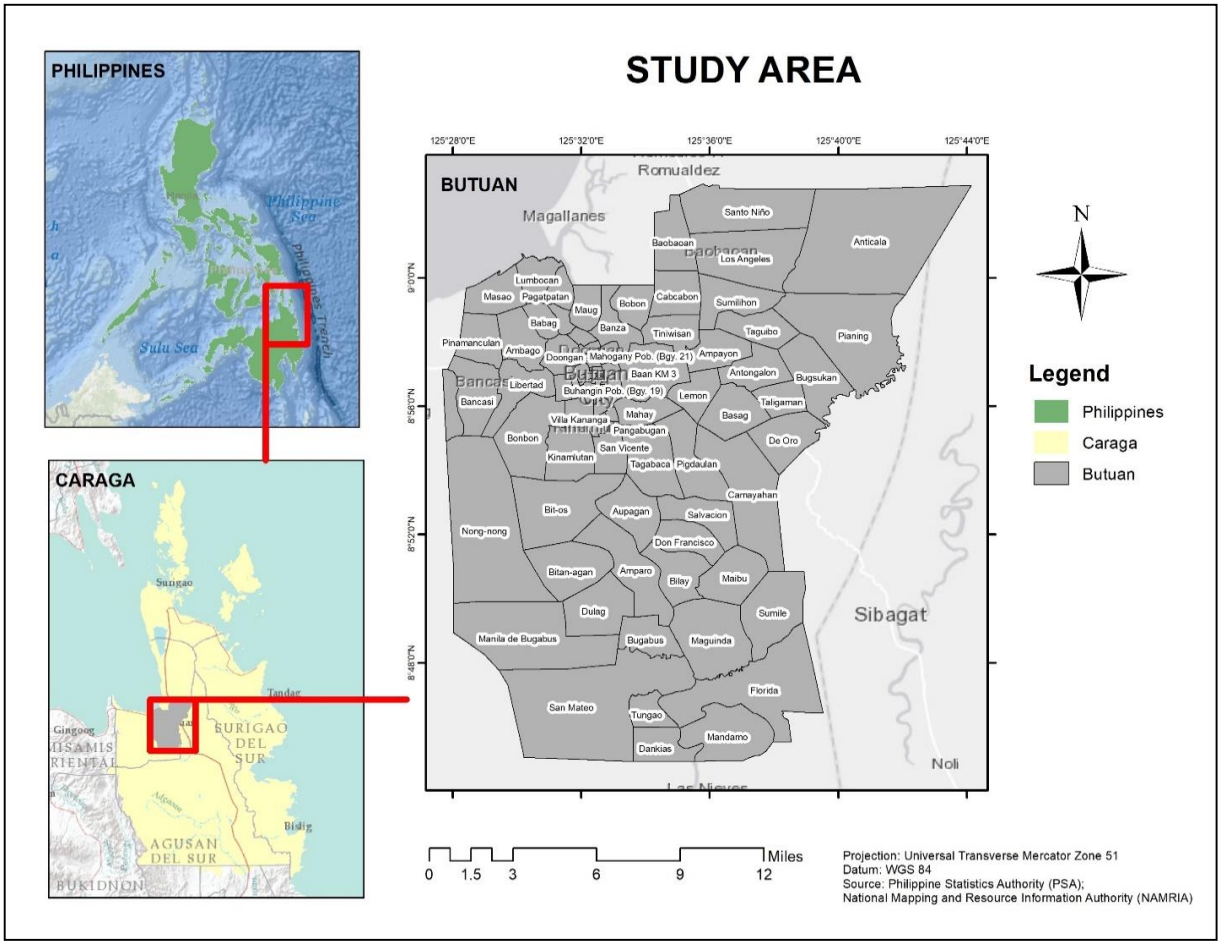

Figure 1. Location Map of Butuan City

2.3 Data Processing

COVID-19 data cases, i.e., confirmed cases, recovered cases, death cases, were collected from the Department of Health (DOH) - Caraga. Confirmed cases and population data per barangay were used in calculating the disease incidence rates. Thematic maps were generated in this study to observe the spatial distribution of COVID-19 incidence in the study region.

2.4 Identifying Neighbors

Figure 2 shows two different contiguities: Queen and Rook contiguity (Renata and Michael, 2005). Primarily encountered in spatial analysis, observations closer together are more similar than at a more extensive distance. It explains that spatial locations are involved in the spread of objects or actions. Queen's case Contiguity was used in this study to identify neighbors, where barangays that share boundaries and corners are considered neighbors (Renata and Michael, 2005).

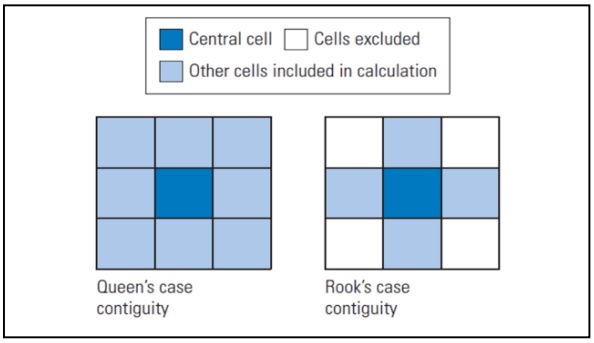

Figure 2. Example of Contiguity Orientation (Queens Case Contiguity and Rook's case Contiguity). Photo Courtesy: Lloyd (2010).

2.5 Global Moran's I

Spatial correlation between variables was assessed using Global Moran's I, which matched location and attribute similarity. Global Moran's Index is calculated based on cross products and mathematically expressed as follows:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(X_i-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})} \quad (1)$$

where *n* denotes the number of barangays; x_i is the attribute value in area I; $\bar{x}$ denotes the attribute mean value in the study area; and W_{ij} denotes the spatial weights of matrix W. The Global Moran's I range from -1 to 1. Positive value indicates that a point is prone to be clustered and dispersed in an opposite value. Values close to zero indicates randomness.

2.6 Getis-Ord General G

A distance-based tool (Getis-Ord General G) was used for calculating the proportion of a variable within a certain radius of a point to the proportion in the entire study region defined as:

$$G_i(d) = \frac{\sum_{j=1}^{n} W_{ij}(d)X_j}{\sum_{i=1}^{n} X_i} \quad (2)$$

where x_j is the value of the observation at point j; $w_{ij}(d)$ is the i and j elements of binary W matrix, w_{ij}=1 if the site is within distance and 0 if elsewhere; n is the number of observations made. The high the z-score indicates the likelihood of high clustering, low z-score for low clustering, while a z-score near to 0 indicating the absence of visible clusters (Li et al., 2020).

2.7 Incremental Spatial Autocorrelation

The Incremental Spatial Autocorrelation tool was used to calculate spatial autocorrelation for a set of distance increments and reports the corresponding Moran's Index, Expected Index, Variance, z-score, and p-value for each distance increment. Z-scores show distances where spatial processes promoting clustering are most prominent, and statistically significant peak z-scores indicate distances where clustering is most pronounced.

2.8 Ripley's K-Function

Ripley's K function was used to see if features, or the values associated with features, are statistically significant at different distances in the study region (Dixon, 2001).

2.9 Local Getis-Ord G_i*

The Local Getis-Ord Gi* was employed in each feature in a dataset. Gz-score and Gp-values classify and identify a feature that exhibits spatial cluster either high or low values. This tool was used to investigate each element in the condition of its surroundings. This tool operates by examining each feature in the context of its surroundings. The

higher the z-score, the stronger the clustering of high values, positive z-score indicates a hot spot. The lower the z-score or the negative z-score indicates a significant cold spot. Getis-Ord Gi* is expressed as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij}x_j - \bar{X}\sum_{j=1}^{n} w_{ij}}{s\sqrt{\frac{[n\sum_{j=1}^{n} w_{ij}^2 - (\sum_{j=1}^{n} w_{ij})^2]}{n-1}}} \quad (3)$$

where x_j is the attribute value for feature j; w_{ij} is the spatial weights between feature i and j; and n is the total number of features (Li et al., 2020).

2.10 Local Moran's I

Cluster and Outlier Analysis finds feature clusters with high or low values. Furthermore, the application detects outliers in space. Local Moran's I analysis (Anselin, 1995) was used to discover statistically significant hot spots, cold spots, and outliers and was mathematically represented as:

$$I_i = X_i \sum_i W_{ij} X_j \quad (4)$$

The Local Moran's I variables and computation are similarly comparable to those of the Global Moran's I where x_i is an attribute for feature I and w_{ij} is the spatial weights between feature i and j. High-High (HH) and Low-Low (LL) are two different types of z-values. High-Low (HL) and Low-High (LH) are the outliers (Li et al., 2020).

3. RESULTS AND DISCUSSION

3.1 Statistical Graph of COVID-19

Figure 3 shows the statistical graph of COVID-19 cases in the study region from Week 1 to Week 45. It can be seen that there was a random trend of COVID-19 confirmed cases each week, and there were increasing trends of cumulative confirmed cases (2,420), cumulative recovered cases (2,324), and cumulative death cases (76) for 45 weeks.

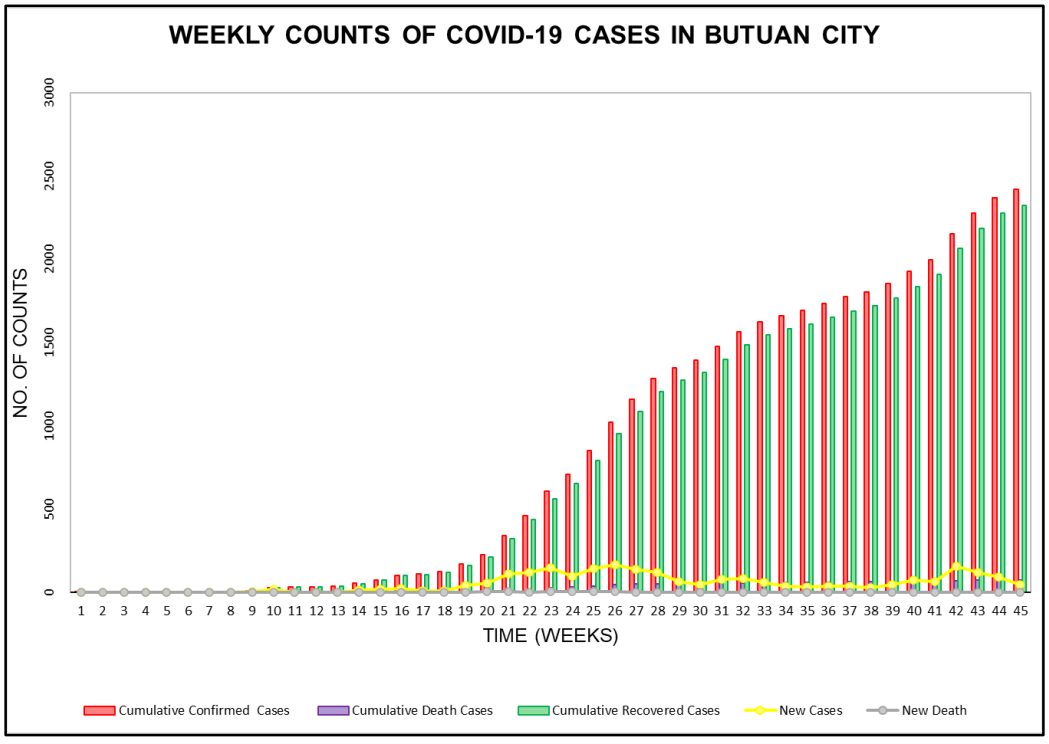

Figure 3. Statistical graph of COVID-19 cases trend

3.2 Thematic Maps of COVID-19 in Butuan City

COVID-19 cases were increasing rapidly in Butuan City, reaching 2,530 as of March 07, 2021. Shown in Figure 4a was the barangay with confirmed cases. Barangays Ambago, Baan Km.3, Doongan, Libertad, San Vicente, and Villa Kananga had the highest confirmed cases and shaded with grey color were the barangays with no confirmed cases (Bugsukan, Camayahan, Salvacion, Bugabus, Dankias, Mandamo). Figure 4b presented the Butuan barangays with recovered cases. Barangays Libertad and Villa Kananga had the highest death cases, as shown in Figure 4c. Moreover, as displayed in Figure 4d, i.e., the static map of disease incidence rate, high incidence areas were in the central part of the city.

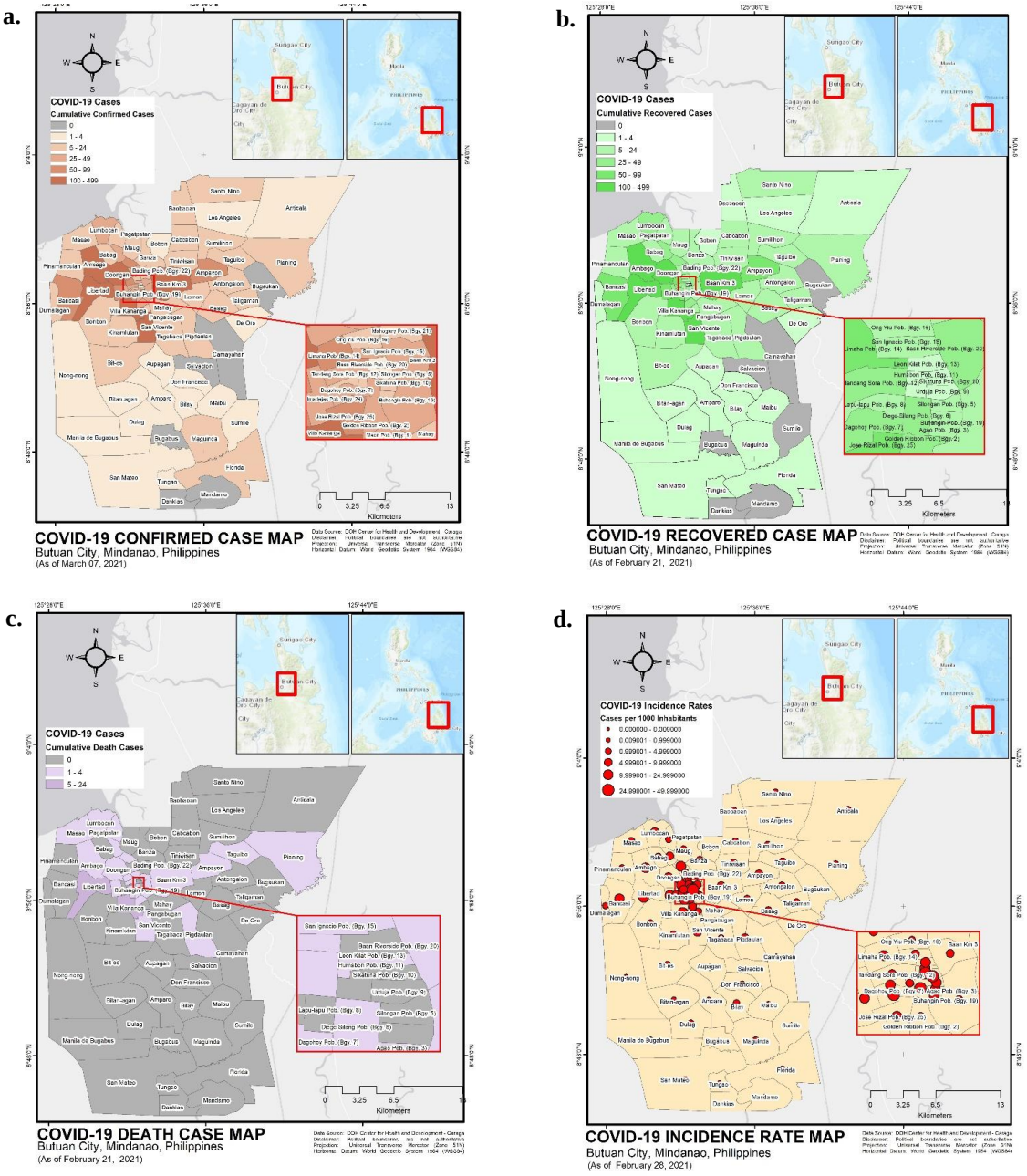

Figure 4. Thematic maps showing the COVID-19 (a) confirmed case, (b) recovered case, (c) death case, and (d) incidence rates.

3.3 Analyzing Patterns

Spatial clusters within the study region must be identified first using global statistics to determine hot spot or cold spot and cluster/outlier of disease rate in the study area. These statistics will indicate clustering and spatial pattern within the study area, and if no presence of clustering, there is no reason to investigate local clusters.

3.3.1 Global Spatial Autocorrelation

Spatial autocorrelation of COVID-19 epidemic between 86 barangays as of June 2021 was lower than zero (negative) according to Global Moran's I calculations as shown in Figure 5a. Indicating randomness or that the incidence of COVID-19 was not spatially correlated between barangays. Global Moran's I for August 2020 cannot identify strong evidence of COVID-19 spatial correlation as presented in Figure 5b. Figure 5c shows strong evidence of clustering, i.e., there is spatial autocorrelation between 86 barangays for February 2021 according to Global Moran's I, z-score, and p-value.

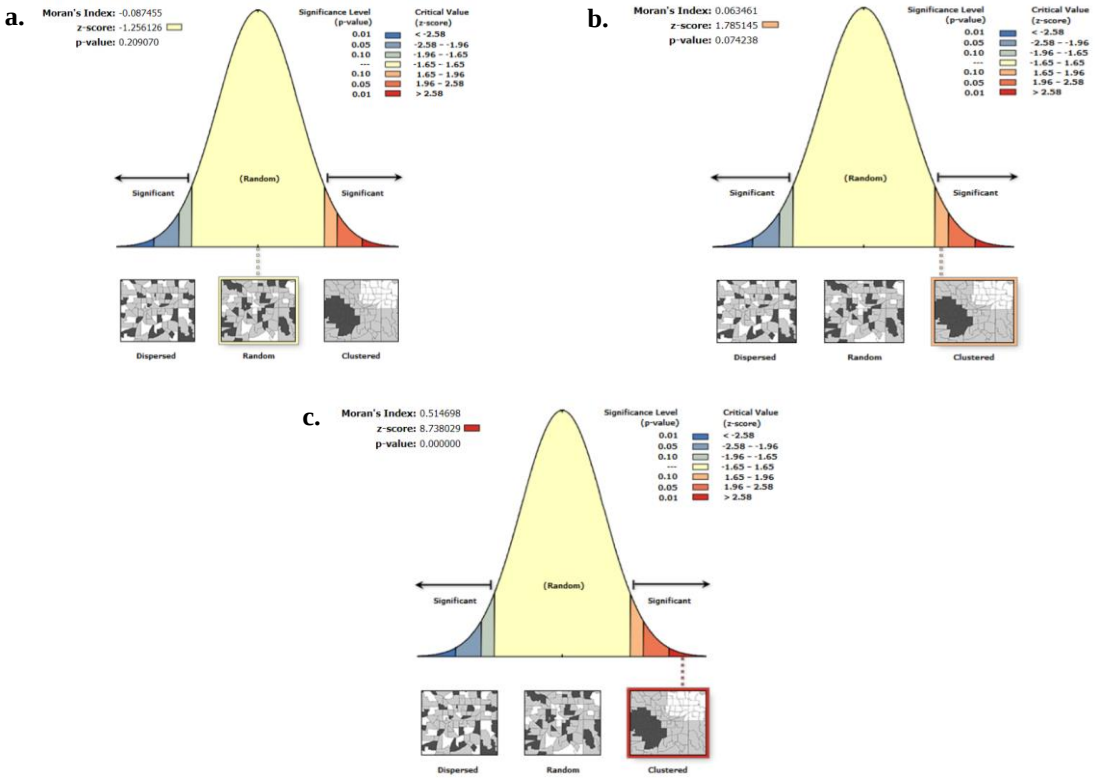

Figure 5. Global Moran's I statistical reports, as of (a) June 2020, (b) August 2020, and (c) February 2021.

3.3.2 High-Low Cluster Analysis

Based on the High-Low Cluster analysis results shown in Figure 6a, the pattern does not appear to be statistically different from random for June 2020. Figure 6b displays the results for August 2020 (General-G=0.107835, z-score 1.857960, p-value=0.063175), indicating a less than 10% likelihood that this high-clustered pattern could be the result of random chance. The high-clustered result for February 2020 shown in Figure 6c (General-G=0.125902, z-score=7.741585, p-value=0.000) indicates there is a less than 1% that this high-clustered could be the result of random chance.

3.3.3 Spatial Autocorrelation by Distance

The spatial autocorrelation for a series of distances using Incremental Spatial Autocorrelation was utilized to determine the spatial clustering. Z-scores reflect the intensity of spatial clustering, and statistically significant peak z-scores indicate distances where spatial processes promoting clustering are most pronounced. As shown in Figure 7a, there is no clustering, z-score falls between -0.1-0.6, while Figures 7b and 7c indicate clustering and are more pronounced on distances 6,400 m and 7,530 m.

3.3.3 Multi-Distance Spatial Cluster Analysis

Determining statistically significant clustering or dispersion over a range of distances is used in this study. When the observed K value is larger than the expected K value at a certain distance, distribution is more clustered than randomness but not statistically significant (see Figure 8a). If the observed K value is smaller than the lower confidence envelope value, spatial dispersion is statistically significant at more extensive distances. Moreover, if the

value of observed K is greater than the high confidence envelope, spatial clustering is statistically significant at smaller distances, as displayed in Figures 8b and 8c.

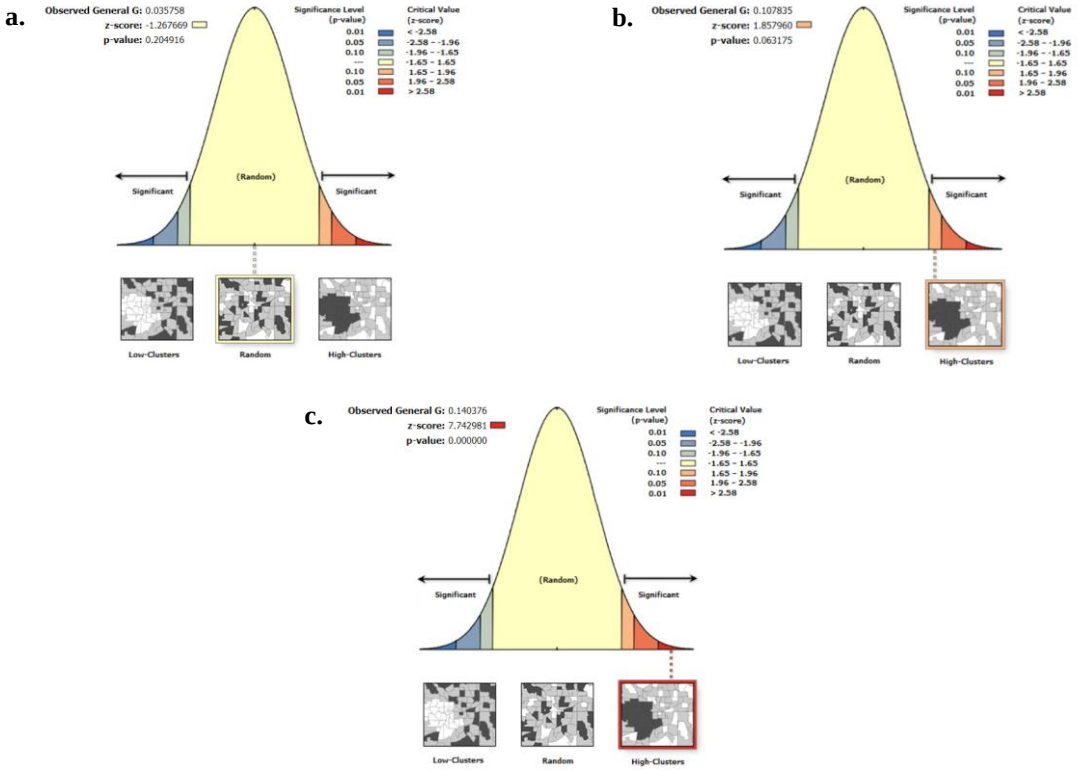

Figure 6. Getis-Ord General G statistical reports, as of (a) June 2020, (b) August 2020, and (c) February 2021.

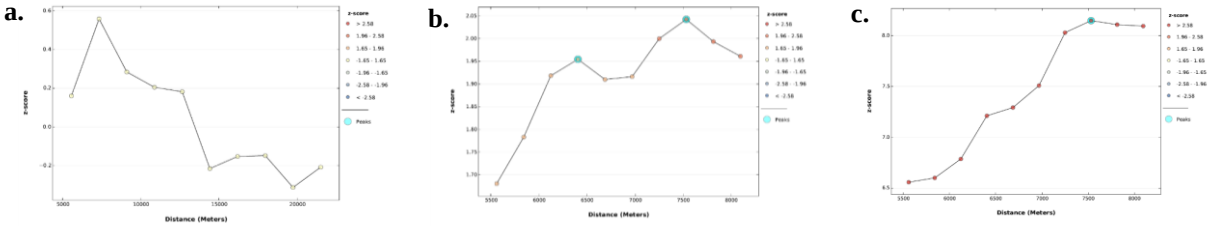

Figure 7. Incremental spatial autocorrelation graph, as of (a) June 2020, (b) August 2020, and (c) February 2021.

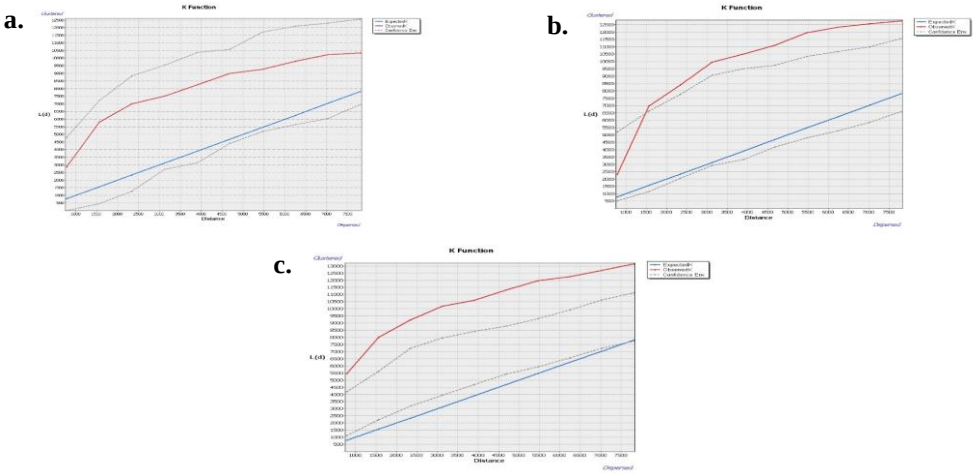

Figure 8. Ripley's K-function statistical graphs, as of (a) June 2020, (b) August 2020, and (c) February 2021.

3.4 Mapping Clusters

Cluster analysis was used to find statistically significant hotspots, cold spots, spatial outliers, and other comparable characteristics or zones.

3.4.1 Hot Spot Analysis

No significant cold spots were present for June 2020. Only four barangays were identified significantly with a 90 – 99% confidence interval shown in Figure 9a. August 2020 displays (HS) barangays with 90 – 99% confidence interval and neighbor barangay with the same confidence interval shown in Figure 9b. Figure 9c shows the result for February, where cold spots and hot spots are both present. Cold spots were clustered at the southern part of the study region, and hot spots were evident at the central portion of the city. The maps revealed that for June to August 2020, only a few barangays were determined as significant. As of February 2021, more barangays were identified significantly, and these are the barangays that exhibit clustering either hotspot or cold spot based on their confidence interval. Table 1 displays the significant barangay with a 90 – 99% confidence interval.

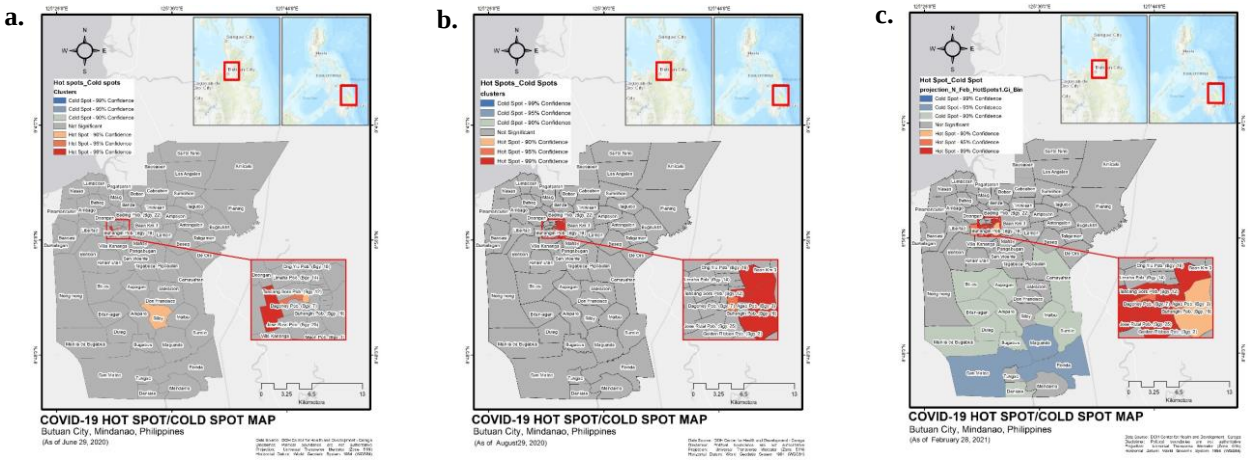

Figure 9. Hot spot analysis maps, as of (a) June 2020, (b) August 2020, and (c) February 2021.

Table 1. Significant hot spot/cold spot barangays

Barangay	GiZ-score	GiP-value	GiBin	Aggregation
Agao	2.160721	0.030717	2	HS-95% CI
Amparo	-1.933714	0.053148	-1	CS-90% CI
Aupagan	-1.6594	0.097035	-1	CS-90% CI
Bayanihan	2.241055	0.025023	2	HS-95% CI
Bit-Os	-1.647669	0.099421	-1	CS-90% CI
Bitan-Agan	-1.761581	0.07814	-1	CS-90% CI
Bugabus	-1.838738	0.065954	-1	CS-90% CI
Buhangin	1.854797	0.063625	1	CS-90% CI
Baan Riverside	3.267107	0.001087	3	HS-99% CI
Imadejas	3.659748	0.000252	3	HS-99% CI
Diego Silang	5.540391	0	3	HS-99% CI
Golden Ribbon	3.54553	0.000392	3	HS-99% CI
Dagohoy	3.514858	0.00044	3	HS-99% CI
JP Rizal	3.92937	0.000085	3	HS-99% CI
Humabon	3.185653	0.001444	3	HS-99% CI
Lapu-Lapu	2.506985	0.012177	2	HS-95% CI
Maguinda	-2.086191	0.036961	-2	HS-95% CI
Maibu	-1.943277	0.051983	-1	CS-90% CI
Manila De Bugabus	-1.85298	0.063885	-1	CS-90% CI
New Society Village	5.13152	0	3	HS-99% CI
Rajah Soliman	2.359746	0.018287	2	HS-95% CI
San Mateo	-2.050832	0.040283	-2	HS-95% CI
Sikatuna	4.924646	0.000001	3	HS-99% CI
Silongan	3.04222	0.002348	3	HS-99% CI
Tandang Sora	4.489493	0.000007	3	HS-99% CI

Urduja	4.678777	0.000003	3	HS-99% CI
Dulag	-1.677569	0.093431	-1	HS-95% CI
Florida	-2.074804	0.038005	-2	CS-95% CI
Camayahan	-1.693003	0.090455	-1	HS-95% CI
Dankias	-1.750231	0.080078	-1	HS-95% CI
Sumile	-1.661475	0.096618	-1	HS-95% CI

3.4.2 Cluster/Outlier Analysis

Only two barangays (HH) were noteworthy (Buhangin and Urduja) for August 2020, as indicated in Figure 10a, and one (LH) cluster barangay (Agao). Five important barangays (HH) are depicted in Figure 10b: (Diego Silang, Dagohoy, Humabon, Sikatuna, and Urduja). Figure 10c is an example map of cluster or outlier analysis demonstrating significant (HH) barangays grouping using the Anselin Local Moran's I. The ten barangays designated as (HH) clusters or high-risk regions are listed in Table 2. From August 2020 to February 2021, no barangay demonstrates High-Low (HL), Low-High (LH), or Low-Low (LL) clustering.

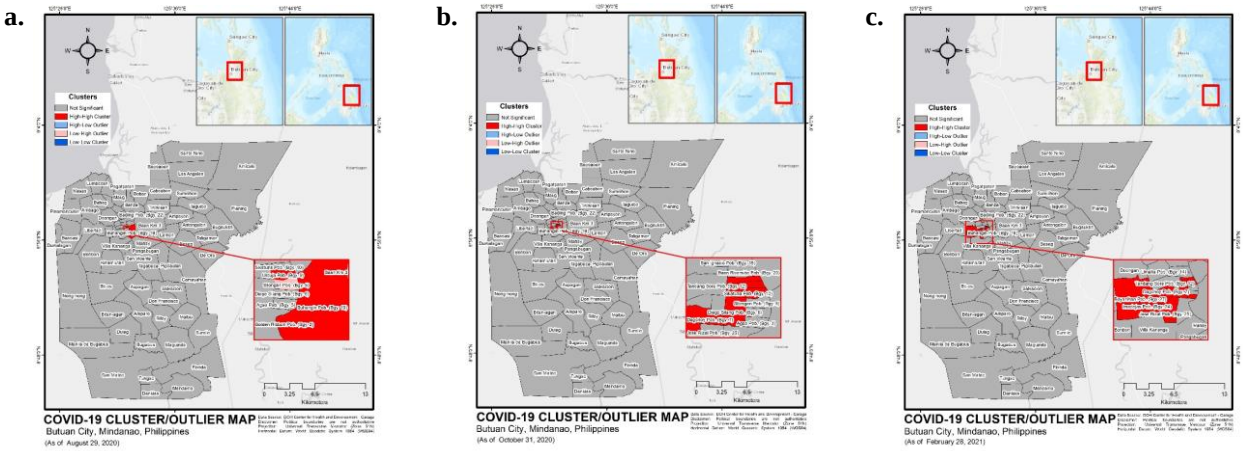

Figure 10. Cluster/outlier analysis maps, as of (a) June 2020, (b) August 2020, and (c) February 2021.

Table 2. Significant high-risk barangays

Barangay	Incidence Rate	LM-Index	LMZ-Score	LMP-Value	Aggregation
Bayanihan	20.004349	5.668429	2.74649	0.006024	HH
Dagohoy	33.135089	17.740196	7.840926	0	HH
Diego Silang	31.938326	40.930032	14.32538	0	HH
Humabon	21.89781	8.889422	4.764718	0.000002	HH
Imadejas	13.306039	6.955371	2.718586	0.006556	HH
JP Rizal	13.105639	5.97862	2.895242	0.003789	HH
Lapu-Lapu	20.264317	5.958502	3.202021	0.001365	HH
Sikatuna	46.511628	36.081038	15.915202	0	HH
Urduja	36.144578	25.792623	12.396998	0	HH
Agao	16.709512	3.974738	2.144326	0.032007	HH

4. CONCLUSION AND RECOMMENDATION

4.1 Conclusion

Based on the results of this study, Moran's I spatial autocorrelation was not observed significant for the first three months from April to July 2020 of COVID-19 incidence in Butuan City. COVID-19 confirmed cases are low for these months, as evident in the thematic maps of confirmed cases. The reported cases are randomly distributed and in different classifications based on their counts—no likelihood of incidence rate in the study region in these months. From August 2020, spatial autocorrelation in the neighbors started to cluster. But this month does not show strong evidence of clustering. It can also result from the random distribution of COVID-19 incidence rate according to the results of Global Moran’s I and Getis-Ord General G. Hotspot analysis was observed monthly in this study. October 2020 began to exhibit cold spots and spread in the southern part of the study region until February 2021. Local spatial autocorrelation was perceived to several barangays, which were identified as significant. Ten barangays were

determined as high-risk areas usually in the urbanized barangays of Butuan City (i.e., Agao, Bayanihan, Dagohoy, Diego Silang, Humabon, Imadejas, JP Rizal, Lapu-Lapu, Sikatuna, Urduja) as of February 2021.

4.2 Recommendation

Data used in this study were only limited to confirmed cases, recovered cases, and death cases. The researchers recommend gathering data that were reported on the onset places and included in the analysis. It is highly suggested to consider other spatial-statistical algorithms to validate the results of spatial autocorrelation (e.g., Bayesian disease mapping, spatial scan statistic). Furthermore, this research can be used in other places where the disease is prevalent.

5. REFERENCES

Dixon, P. M., 2001. Ripley's K function Theoretical K(t) function. Statistics (Ber)., vol. 3, no. December, pp. 1–16

Kang, D., Choi, H., Kim, J. H., and Choi, J., 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. Int. J. Infect. Dis., vol. 94, no. January, pp. 96–102.

Li, H., Li, H., Ding, Z., Hu, Z., Chen, F., Wang, K., Peng, Z., and Shen, H., 2020. Spatial statistical analysis of coronavirus disease 2019 (Covid-19) in China. Geospat. Health, vol. 15, no. 1, pp. 11–18, 2020.

Lloyd, C., 2010. Spatial data analysis: an introduction for GIS users. Oxford University Press. ISBN 978-0-19-955432-4

Ramirez-Aldana, R., Gomez-Verjan, J. C., and Bello-Chavolla, O. Y., 2020. Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level.

Renata, K., and Michael, R., 2005. Spatial data analysis in R, vol. 2005-January.

Shariati, M., Mesgari, T., Kasraee, M., and Jahangiri-rad, M., 2020. Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system (March and April, 2020). J. Environ. Heal. Sci. Eng.

Xiong, Y., Wang, Y., Chen, F., and Zhu, M., 2020. Spatial statistics and influencing factors of the COVID-19 epidemic at both prefecture and county levels in Hubei Province, China. Int. J. Environ. Res. Public Health, vol. 17, no. 11.