VOL 4, NO.2 (2024) 169-186
https://jagst.utm.my

JAGST
Journal of Advanced Geospatial Science & Technology

Hotspot Analysis of COVID-19 using Spatial Statistic in Selangor Malaysia

Nurain Othman, Zamri Ismail*
Geoinformation Program, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

*Corresponding author: zamriismail@utm.my

Abstract **–** The study investigates the spatial distribution of COVID-19 cases in Selangor, Malaysia, utilizing geospatial and geostatistical techniques to identify and analyze hotspots. Focusing on data from 54 districts between April and August 2021, the research employs Global Moran's I and Getis-Ord Gi* models to detect spatial autocorrelation and clustering patterns. The study also incorporates additional parameters, including new cases, cumulative cases, deaths, clusters, and population density, to comprehensively analyze COVID-19 hotspots. The results reveal a significant clustering of COVID-19 cases, with specific districts like Petaling, Hulu Langat, and Klang identified as high-risk hotspots. The findings of this research emphasize the critical role of spatial analysis in understanding the spread of infectious diseases like COVID-19. By identifying and mapping out high-risk districts, this study provides valuable insights that can inform public health strategies and optimize resource allocation in response to the pandemic. Identifying hotspots within Selangor underscores the necessity for targeted interventions and deploying healthcare resources to areas most affected by the virus. Ultimately, this study contributes to a deeper understanding of the spatial dynamics of COVID-19 in Selangor, offering a framework for future research and public health planning in the context of epidemic management.

Keywords *– COVID-19, Geospatial Analysis, Global Moran's I, Getis-Ord Gi*, Hotspots*

Article History: Received 26 March 2024, Accepted 8 August 2024, Published 31 August 2024

How to cite: Othman, N. and Ismail, Z. (2024). Hotspot Analysis of COVID-19 Using Spatial Statistic Technique in Selangor State. Journal of Advanced Geospatial Science and Technology. 4(2), 169-186.

1.0 Introduction

1.1 Background

A novel coronavirus was discovered as the source of an infection epidemic in China in 2019. This novel coronavirus developed in Hubei Province, China, and the infection by the virus is called COVID-19 (Surveillances, 2020). The first cases in Malaysia were recorded in Johor Bahru, involving a 65-year-old lady and two children, ages two and eleven. Then, the cases were restricted to a few imported cases until March 2020, when many local clusters had since developed. Since then, the number of positive cases in Malaysia has risen (Fatima et al., 2021). Selangor was reported as the state with the most COVID-19-positive cases compared with other states in Malaysia (Hakim, 2021).

A hotspot is a geospatial technique that assures cases are located in the centre. Much research has used variable techniques to analyze these hotspots, which may help distinguish between the hotspots and normal regions (Xu et al., 2021). Those methods include Getis–Ord Gi*, Kernel Density Estimation, Nearest Neighbour Index, and Standard Deviational Ellipses.

This research will utilize the number of new cases, cumulative cases, deaths, cluster, and population density to map the Selangor state's hotspot detection. Many nations have been attempting to identify COVID-19 hotspots to take appropriate action, depending on confirmed and positive cases that have been documented. The current study, on the other hand, employed geospatial and geostatistical techniques, as well as other important parameters rather than just COVID-19 cases, to find the most critical component that may be considered for detecting hotspots of new coronavirus development in Selangor. This included research in 54 districts in Selangor for five (5) months, from April to August 2021.

The objectives of the research are: (1) to determine the parameter for the hotspot analysis; (2) to develop a geospatial database for the hotspot analysis; and (3) to generate a hotspot map and to analyze the factor contributing to higher COVID-19 deaths. Since this research mainly focuses on hotspot detection of COVID-19, the significance of this research on hotspot analysis can help to recognize which places in a study area should receive more attention, which locations are more impacted, and which areas are highly susceptible soon.

2.0 Literature Review

2.1 Hotspot Detection Using Spatial Statistics

According to Columbia Public Health, hotspot analysis is a spatial analysis and mapping technique interested in identifying and clustering spatial phenomena. These spatial phenomena are depicted as points on a map and refer to locations of events or objects (Columbia Public Health, 2023). Spatial autocorrelation and cluster analysis are two approaches for analyzing spatial patterns and detecting hotspots.

The spatial autocorrelation study examines how effectively items correlate with neighbouring objects across a geographical region (Getis & Ord, 2008). Positive autocorrelation occurs when numerous identical values are discovered close together, and negative correlation occurs when significantly diverse outcomes are found close together. The significance of spatial autocorrelation is that it helps to determine how essential spatial qualities are in impacting a specific item in space and whether there is a clear link between objects and spatial attributes (Cliff & Ord, 1981). Significantly negative or positive findings suggest that the object has a distinct spatial feature with a high correlation.

The idea of cluster analysis is to identify natural object segmentation. In other words, it divides related observations into homogenous subsets. These subclasses might indicate patterns linked to the investigated issue (Everitt et al., 2011). A distance approach is utilized to establish whether units are similar, and various clustering algorithms based on multiple conceptions are provided. When data is clustered, similarity measures are taken between the data and then between the clusters.

Hotspot analysis techniques, such as Getis-Ord Gi* and Local Moran's I, are commonly used in epidemiological studies to identify and visualize the spatial distribution of diseases, including COVID-19 (Anselin, 1995; Getis & Ord, 1996). These techniques allow researchers to identify areas with a high concentration of cases (hotspots) and areas with a low concentration of cases (cold spots). Understanding these patterns is crucial for public health interventions and resource allocation.

2.2 Models

Throughout this study, the models used are Global Moran's I and Getis-Ord Gi*. The degree of resemblance about other surrounding items was determined using spatial autocorrelation. Global

Moran's I statistics were used to quantify spatial correlation in general, with three categories of categorization statistics: positive, negative, and no autocorrelation (Moran, 1948). To distinguish between hotspots and cold spots, Getis-Ord Gi* will be utilized (Getis & Ord, 2008).

The degree of the selected characteristics' distribution patterns, such as cluster, scattered, and random over the research region, is measured using Global Autocorrelation or Global Moran's I, which Patrick Moran proposed in 1948 (Moran, 1948). The statistical significance of the Global Autocorrelation is determined by comparing all of the specified characteristics by locations and attributes. A positive Moran's I index value shows a trend toward clustering when the z-score or p-value implies statistical significance. In contrast, a negative Moran's I index value shows a trend toward dispersion (Cliff & Ord, 1981). This tool computes a z-score and p-value to determine whether or not the null hypothesis could be rejected. The null hypothesis in this example asserts that feature values are distributed randomly over the research region. The Input Field must be filled with a variety of values. This statistic's calculation requires some variance in the evaluated parameter (Getis & Ord, 2008).

To determine the spatial relationship between the feature's high and low values, local G statistics or Getis-Ord Gi* are utilized (Getis & Ord, 2008). The output that the Getis-Ord Gi* provides is a z-score, which, unlike Global and Local Moran's I, does not require further Z-score calculation (Monzur, 2019). The z-scores and p-values are statistical significance measures that tell if the null hypothesis should be rejected feature by feature. In sum, they show that if the observed spatial clustering of high or low values is more robust than one would anticipate from a random distribution of those same values, The False Discovery Rate, known as FDR adjustment, is not applied to the z-score and p-value variables. Whether or not the FDR correction is done, the Gi Bin field indicates statistically significant hot and cold regions (Fotheringham et al., 2002). Features in the +/−3 bins have a 99 per cent confidence level of statistical significance; features in the +/−2 bins have a 95 per cent confidence level; features in the +/−1 bins have a 90 per cent confidence level; and features in bin 0 have no statistical relevance (Fotheringham et al., 2002).

The rationale behind selecting Global Moran's I and Getis-Ord Gi* lies in their ability to identify and differentiate clusters and hotspots within spatial data effectively. Global Moran's I is suitable for detecting overall spatial patterns, while Getis-Ord Gi* is advantageous for identifying localized clusters. These techniques offer a comprehensive understanding of spatial distribution patterns, which is crucial for analyzing COVID-19 data (Getis & Ord, 2008; Moran, 1948).

2.3 Previous Study on Hotspot Detection

This section summarizes articles based on previous studies on the Hotspot Detection of COVID-19. These articles employed different techniques and approaches for data processing and data analysis. These studies provide good references in general for producing a good hotspot detection of COVID-19 in Selangor.

A study examined the spatial distribution of COVID-19 crude rates on a global scale through the application of global and local spatial autocorrelation techniques of cluster detection and hotspot analysis to provide the geospatial analysis of the COVID-19 pandemic for the year 2020 (Su et al., 2020). The data was processed using Getis-Ord Gi* for hotspot analysis, Anselin Local Moran's I for spatial cluster outlier, and Global Moran's I for global autocorrelation.

Another research utilized various spatial analysis methods to identify and analyze COVID-19 hotspots in multiple regions (Rex et al., 2020). These methods included Kernel Density Estimation (KDE) to create a continuous surface density of COVID-19 cases, which helped visualize areas with a high concentration of cases over a continuous landscape. The Nearest Neighbour Index (NNI) was also used to determine the clustering tendency of COVID-19 cases by comparing the observed average distance between cases to the expected average distance in a random distribution. These methods provided a more nuanced understanding of the spatial distribution and clustering patterns of COVID-19 cases.

Using Standard Deviational Ellipses (SDE) in hotspot detection helped identify the directional distribution of COVID-19 cases, which can be crucial for understanding the spread dynamics of the virus (Li et al., 2020). This method provided additional insights into the spatial distribution patterns by summarizing the spatial characteristics of the case locations, including the central tendency, dispersion, and directional trends.

Overall, combining these various spatial analysis methods and models has proven effective in identifying and understanding the spatial distribution patterns of COVID-19 cases. These techniques provide valuable insights that can inform public health strategies and interventions to mitigate the spread of the virus.

3.0 Methodology

The research methodology has been produced to achieve all three research objectives. This study is divided into four phases of Methodology: Phase 1 focuses mainly on the preliminary research and literature review, as discussed in Section 2.0, whereas Phase 2 highlights data acquisition; Phase 3 on data processing; and lastly, Phase 4 on results and analysis.

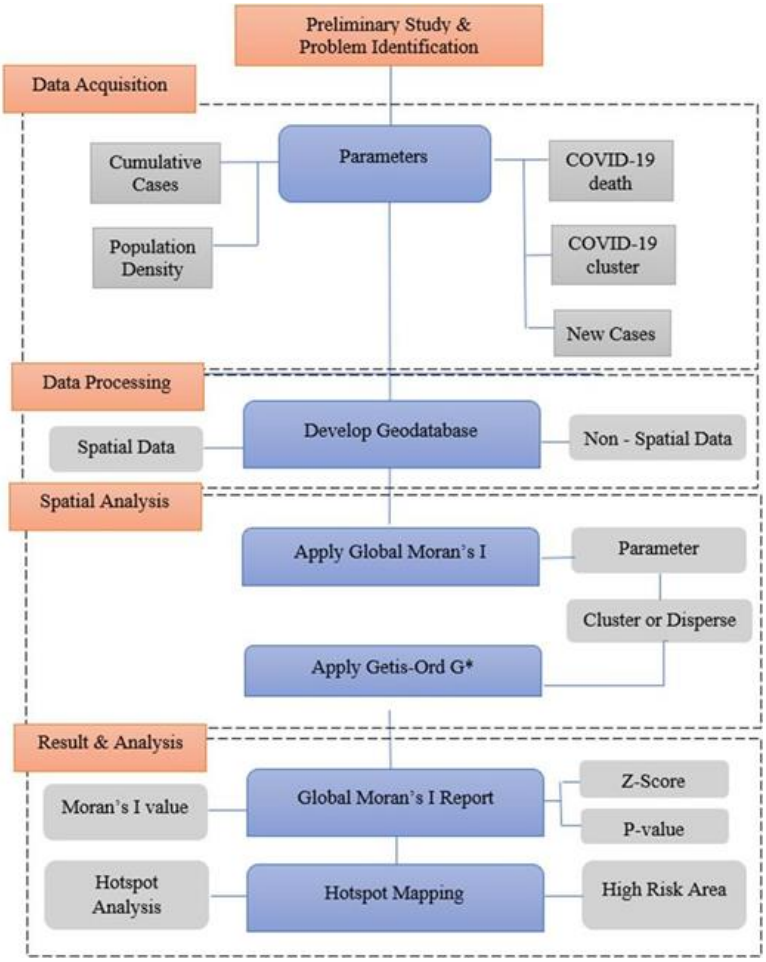


Figure 1. Research method

3.1 Data Acquisition

Data acquisition is a crucial phase that must be completed once the preliminary research and literature evaluation has been completed. This phase will elucidate the source data: number of new cases, number of cumulative cases, number of deaths, population density, and number of clusters.

The data related to COVID-19 cases, which are number of deaths by COVID-19, cumulative cases, new cases, and number of COVID-19 clusters from April 2021 until August

2021, are collected from the Ministry of Health's official website. As for the population density data, the population in Selangor is collected from the Department of Statistics Malaysia for 2020. Other than that, the Selangor Boundary of spatial data was also used to develop the Mukim Boundary.

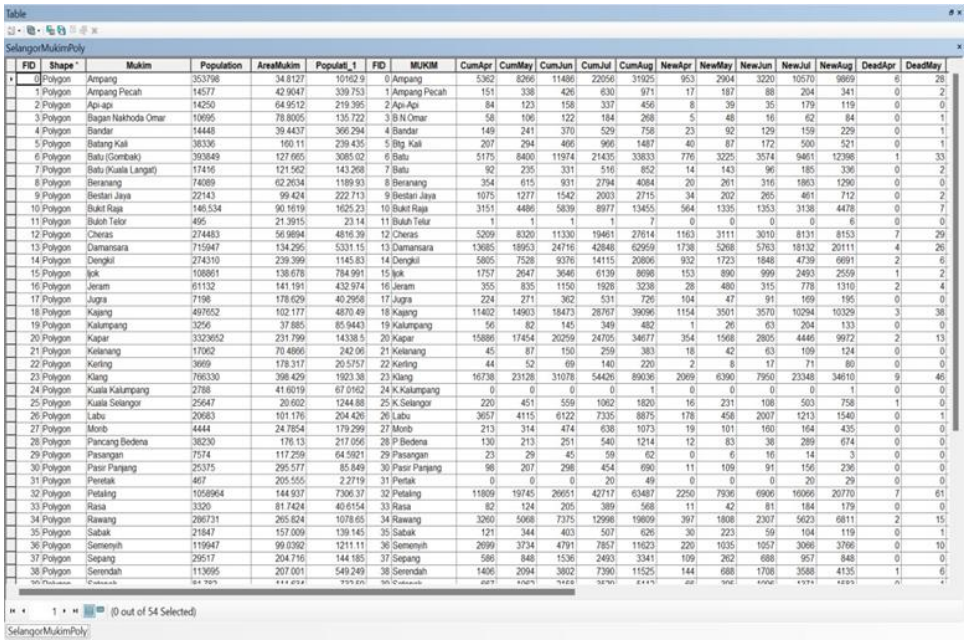
3.2 Data Processing

3.2.1 Develop Geospatial Database

Data processing is obtaining raw data and transforming it into usable data. Firstly, the Geospatial Database was developed in ArcMap 10.4 before it could begin to store the spatial and non-spatial data. The geodatabase is created by creating a File Geodatabase from the folder Connection in ArcMap.

3.2.2 Processing of Spatial and Non-Spatial Data

The spatial data of Mukim Boundary were developed using Selangor Boundary data by georeferencing the image of the Selangor district. The Excel spreadsheets of the non-spatial data are joined into the district point attribute for data processing. To get the population density data, the area of each district is calculated on ArcMap using the Calculate Geometry function. Then, population density is calculated using the Field Calculator function by utilizing population and area of district data. After all of the non-spatial data are completed in ArcMap (Figure 2), the point feature of the Mukim Boundary was generated as in Figure 3 to kickstart the analysis.



Table

SelangorMukimPoly

FID	Shape *	Mukim	Population	AreaMukim	Populati_1	FID	MUKIM	CumApr	CumMay	CumJun	CumJul	CumAug	NewApr	NewMay	NewJun	NewJul	NewAug	DeadApr	DeadMay
0	Polygon	Ampang	353798	34.8127	10162.9	0	Ampang	5362	8266	11486	22056	31925	953	2904	3220	10570	9869	6	26
1	Polygon	Ampang Pecah	14577	42.9047	339.753	1	Ampang Pecah	151	338	426	630	971	17	187	88	204	341	0	2
2	Polygon	Api-api	14250	64.9512	219.395	2	Api-Api	84	123	158	337	456	8	39	35	179	119	0	0
3	Polygon	Bagan Nakhoda Omar	10695	78.8005	135.722	3	B.N Omar	58	106	122	184	268	5	48	16	62	84	0	1
4	Polygon	Bandar	14448	39.4437	366.294	4	Bandar	149	241	370	529	758	23	92	129	159	229	0	1
5	Polygon	Batang Kali	38336	160.11	239.435	5	Btg. Kali	207	294	466	966	1487	40	87	172	500	521	0	1
6	Polygon	Batu (Gombak)	393849	127.665	3085.02	6	Batu	5175	8400	11974	21435	33833	776	3225	3574	9461	12398	1	33
7	Polygon	Batu (Kuala Langat)	17416	121.562	143.268	7	Batu	92	235	331	516	852	14	143	96	185	336	0	2
8	Polygon	Beranang	74089	62.2634	1189.93	8	Beranang	354	615	931	2794	4084	20	261	316	1863	1290	0	0
9	Polygon	Bestari Jaya	22143	99.424	222.713	9	Bestari Jaya	1075	1277	1542	2003	2715	34	202	265	461	712	0	2
10	Polygon	Bukit Raja	146.534	90.1619	1625.23	10	Bukit Raja	3151	4486	5839	8977	13455	564	1335	1353	3138	4478	0	7
11	Polygon	Buloh Telor	495	21.3915	23.14	11	Buluh Telur	1	1	1	1	7	0	0	0	0	6	0	0
12	Polygon	Cheras	274483	56.9894	4816.39	12	Cheras	5209	8320	11330	19461	27614	1163	3111	3010	8131	8153	7	29
13	Polygon	Damansara	715947	134.295	5331.15	13	Damansara	13685	18953	24716	42848	62959	1738	5268	5763	18132	20111	4	26
14	Polygon	Dengkil	274310	239.399	1145.83	14	Dengkil	5805	7528	9376	14115	20806	932	1723	1848	4739	6691	2	6
15	Polygon	Ijok	108861	138.678	784.991	15	Ijok	1757	2647	3646	6139	8698	153	890	999	2493	2559	1	2
16	Polygon	Jeram	61132	141.191	432.974	16	Jeram	355	835	1150	1928	3238	28	480	315	778	1310	2	4
17	Polygon	Jugra	7198	178.629	40.2958	17	Jugra	224	271	362	531	726	104	47	91	169	195	0	0
18	Polygon	Kajang	497652	102.177	4870.49	18	Kajang	11402	14903	18473	28767	39096	1154	3501	3570	10294	10329	3	38
19	Polygon	Kalumpang	3256	37.885	85.9443	19	Kalumpang	56	82	145	349	482	1	26	63	204	133	0	0
20	Polygon	Kapar	3323652	231.799	14338.5	20	Kapar	15886	17454	20259	24705	34677	354	1568	2805	4446	9972	2	13
21	Polygon	Kelanang	17062	70.4866	242.06	21	Kelanang	45	87	150	259	383	18	42	63	109	124	0	0
22	Polygon	Kerling	3669	178.317	20.5757	22	Kerling	44	52	69	140	220	2	8	17	71	80	0	0
23	Polygon	Klang	768330	398.429	1923.38	23	Klang	16738	23128	31078	54426	89036	2069	6390	7950	23348	34610	9	46
24	Polygon	Kuala Kalumpang	2788	41.6019	67.0162	24	K.Kalumpang	0	0	0	0	1	0	0	0	0	1	0	0
25	Polygon	Kuala Selangor	25647	20.602	1244.88	25	K.Selangor	220	451	559	1062	1820	16	231	108	503	758	1	0
26	Polygon	Labu	20683	101.176	204.426	26	Labu	3657	4115	6122	7335	8875	178	458	2007	1213	1540	0	1
27	Polygon	Morib	4444	24.7854	179.299	27	Morib	213	314	474	638	1073	19	101	160	164	435	0	0
28	Polygon	Pancang Bedena	38230	176.13	217.056	28	P.Bedena	130	213	251	540	1214	12	83	38	289	674	0	0
29	Polygon	Pasangan	7574	117.259	64.5921	29	Pasangan	23	29	45	59	62	0	6	16	14	3	0	0
30	Polygon	Pasir Panjang	25375	295.577	85.849	30	Pasir Panjang	98	207	298	454	690	11	109	91	156	236	0	0
31	Polygon	Peretak	467	205.555	2.2719	31	Pertak	0	0	0	20	49	0	0	0	20	29	0	0
32	Polygon	Petaling	1058964	144.937	7306.37	32	Petaling	11809	19745	26651	42717	63487	2250	7936	6906	16066	20770	7	61
33	Polygon	Rasa	3320	81.7424	40.6154	33	Rasa	82	124	205	389	568	11	42	81	184	179	0	0
34	Polygon	Rawang	266731	265.824	1078.65	34	Rawang	3260	5068	7375	12998	19809	397	1808	2307	5623	6811	2	15
35	Polygon	Sabak	21847	157.009	139.145	35	Sabak	121	344	403	507	626	30	223	59	104	119	0	1
36	Polygon	Semenyih	119947	99.0392	1211.11	36	Semenyih	2699	3734	4791	7857	11623	220	1035	1057	3066	3766	0	10
37	Polygon	Sepang	29517	204.716	144.185	37	Sepang	586	848	1536	2493	3341	109	262	688	957	848	0	0
38	Polygon	Serendah	113695	207.001	549.249	38	Serendah	1406	2094	3802	7390	11525	144	688	1708	3588	4135	1	6

1 (0 out of 54 Selected)

SelangorMukimPoly

Figure 2. Attribute data used for data processing

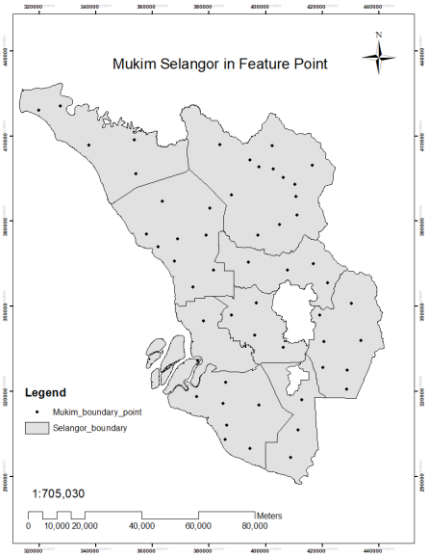


Figure 3. Mukim Boundary in Point Feature

3.2.3 Apply Global Moran's I

Hotspot analysis depends on the availability of spatial clustering in the data. Hence, Global Moran's I will be applied in this study using ArcMap 10.4. Global Moran's I is a tool that determines spatial autocorrelation by employing feature locations and attribute values. Before applying Moran's I, the Conceptualisation of Spatial Relationships, the parameter will be chosen

to represent the fundamental links between the features for analysis. The fixed band distance method has been selected, which works well with point data.

The Global Moran's I tool produces five key results: the Moran's I Index, Expected Index, Variance, Z-score, and P-value. The Moran's I Index ranges from -1 to +1, where +1 indicates perfect clustering, 0 indicates a random spatial pattern, and -1 indicates ideal dispersion. The Expected Index is the value of Moran's I under the null hypothesis of no spatial autocorrelation, serving as a benchmark to determine if the observed Moran's I index significantly deviates from random distribution. The Variance indicates the variability of the Moran's I index, helping to determine the statistical significance of the observed spatial autocorrelation. The Z-score shows how many standard deviations the observed Moran's I index is from the expected index, with a high positive Z-score suggesting clustering and a low negative Z-score suggesting dispersion. The P-value indicates the probability that the observed spatial pattern is due to chance, with a low P-value (typically < 0.05) indicating a statistically significant pattern.

These values are accessible via the Results tab and provided as output values for use in models or scripts. Additionally, this tool can generate an HTML file with a visual report of the results (Figure 4). For example, an output might show a Moran's I Index of 0.069970, an Expected Index of -0.018868, a Variance of 0.000491, a Z-score of 4.009499, and a P-value of 0.000061, indicating a significant clustering of COVID-19 cases in the study area.

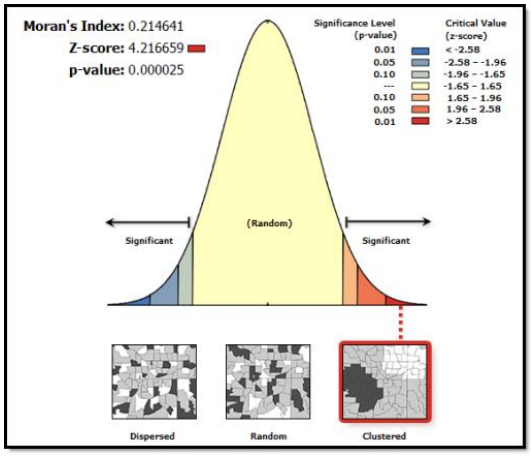


Figure 4. Example of spatial autocorrelation report

*3.2.4 Apply Getis–Ord Gi**

Getis–Ord Gi* statistics will be applied to obtain information on Selangor's COVID-19 hotspots and cold spots. Getis–Ord Gi* generates a new Output Feature Class for each feature in the Input Feature Class, complete with a Z-score, P-value, and confidence level bin (Gi Bin). Statistical significance is determined using the P-value and Z-score fields without FDR correction. If the FDR Correction parameter is applied, the essential P-values defining confidence levels are decreased. The fixed band distance method is used with a threshold distance of 50 kilometres. When the Output Feature Class of Hotspot Analysis is generated, the hotspot mapping for each parameter will be created, allowing the pattern of hotspots to be observed via ArcMap 10.4.

The outputs include the Gi Z-Score, P-value, and Gi Bin. The Gi Z-Score measures the intensity of clustering, with a high positive Z-score indicating a hotspot and a high negative Z-score indicating a cold spot. The P-value indicates the statistical significance of the Z-score, with a low P-value indicating that the hotspot or cold spot is statistically significant. The Gi Bin categorizes the confidence level of each feature into bins based on the Z-score and P-value, showing statistical significance (e.g., 90%, 95%, 99%).

4.0 Results and Analysis

4.1 Spatial Autocorrelation

Table 1 shows the results of Global spatial autocorrelation at a fixed distance band of 50 kilometres. The global spatial autocorrelation showed that all the parameters were clustered, resulting from Moran's I value being higher than 0, and the results were also highly significant, with the value of p being less than 0.01. Therefore, the null hypothesis is rejected since all the z-scores are greater than 1.96, indicating that all the parameters are highly significant. Global Moran's I reported that the cumulative cases generated the highest Moran's I value of 0.123709, resulting from the highest critical value (z-score) of 6.057562. There is less than 1% likelihood that this clustered pattern could result from random chance. On the other hand, population density produces the lowest value of Moran's Index, which produces the lowest value of z-score of 4.009499.

Table 1. Global spatial autocorrelation analysis of COVID-19 by using Global Moran's I

Parameters	Moran's Index	Expected Index	Variance	Z-score	P-value	Pattern
Population Density	0.069970	−0.018868	0.000491	4.009499	0.000061	Clustered
New Cases	0.109685	−0.018868	0.000537	5.549976	0.000000	Clustered
Cumulative Cases	0.123709	−0.018868	0.000554	6.057562	0.000000	Clustered
COVID-19 Deaths	0.099257	−0.018868	0.000555	5.015994	0.000001	Clustered
COVID-19 Clusters	0.096285	−0.018868	0.000524	5.029043	0.000000	Clustered

4.2 Hotspot Analysis

The hotspot mapping for each parameter, new cases, cumulative cases, population density, COVID-19 clusters, and COVID-19 deaths, is being produced. Ten hotspot maps are being produced. The hotspot mapping based on new cases in Figure 6, the cold spot area focuses on the Ulu Bernam, Sungai Panjang, Pasir Panjang, Pancang Bedena, Tanjung Karang, Sungai Tinggi, Peretak, Kerling, Ampang Pechah, Sungai Gumut, Kalumpang, and Kuala Kalumpang with the value of confidence level is 90%and 95% 99%. On the contrary, Ulu Langat, Ampang, Ulu Semenyih, Cheras, Kajang, Semenyih, Beranang, Dengkil, Labu, Sepang, Klang, Kapar, Tanjong Dua Belas, Telok Panglima Garang, Bandar, Jugra, Kelanang, Morib, Batu (Kuala Langat), Ulu Kelang, Rawang, Batu (Gombak), Sungai Buloh, Bukit Raja, Damansara, and Petaling are detected as hotspot 90%, 95% and 99% confidence level.

On the other hand, the hotspot area for cumulative cases in Figure 7 has the same trend as a hotspot of new cases but is slightly different in the cold spot area. Compared with new cases, Mukim Rasa has become a cold spot with a confidence level of 90%, which is the low value of z-score but higher than 1.96. Sungai Gumut stayed in a cold spot but had a higher confidence level of 99%.

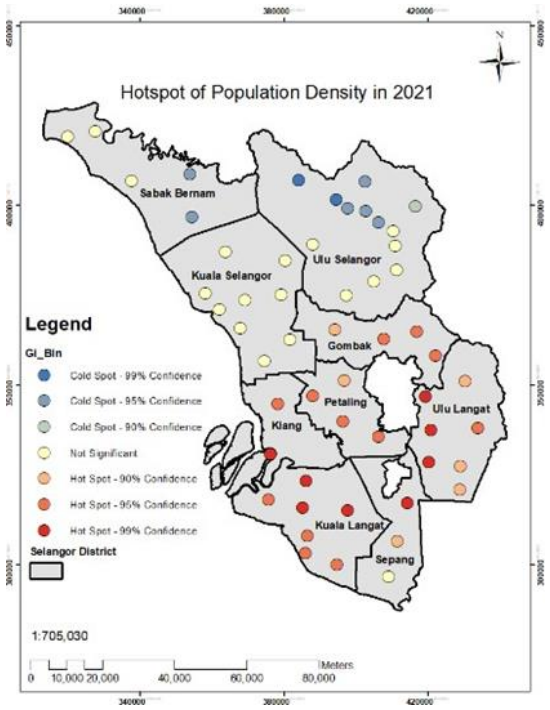


Figure 5. Hotspot detection of population density in 2021

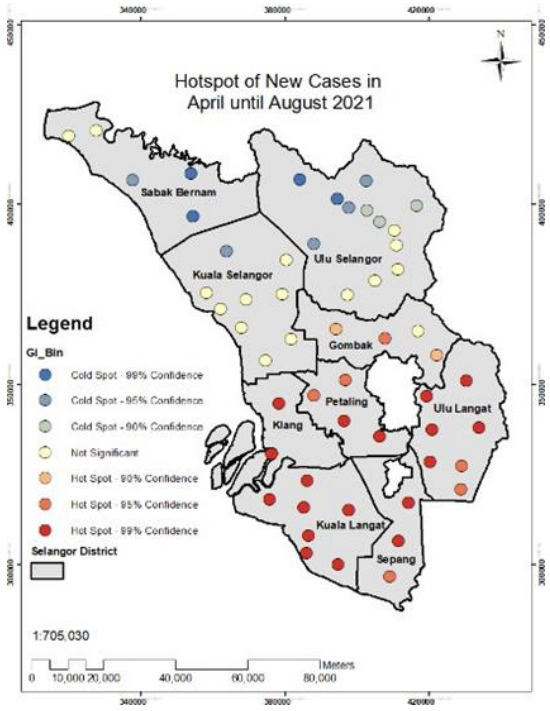


Figure 6. Hotspot detection of new cases from April to August 2021

Based on Figure 5 on population density, Klang, Kapar, Sungai Buloh, Bukit Raja, Damansara, Petaling, Jugra, Telok Panglima Garang, Bandar, Kelanang, Morib, Batu, Tanjung Dua Belas, Labu, Dengkil, Kajang, Beranang, Semenyih, Cheras, Ulu Semenyih, Ulu Langat, Ampang, Ulu Kelang, Setapak, Batu, and Rawang had shown a significant hotspot with confidence level of 90%, 95% and 99%. Half are in Sabak Bernam and Ulu Selangor, which are Sungai Panjang, Pasir Panjang, Ulu Bernam, Kuala Kalumpang, Kalumpang, Sungai Gumut, Kerling, Peretak, and Ampang Pechah are detected as a cold spot with a confidence level of 90%, 95% and 99%, hence the other area is identified as not significant.

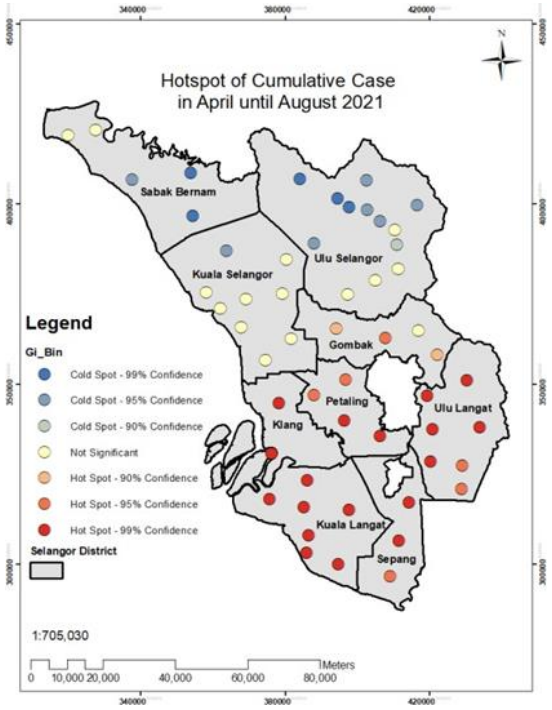


Figure 7. Hotspot detection of cumulative cases from April to August 2021

Cold spots of COVID-19 clusters (Figure 8) are in the majority area of Ulu Selangor and Sabak Bernam, with the addition of two mukim in Kuala Selangor. The cold spot area in those places is 90% and 95% confidence level. Meanwhile, the hotspot area is focusing on Kapar, Klang, Damansara, Petaling, Jugra, Telok Panglima Garang, Bandar, Ketanang, Morib, Batu (Kuala Langat), Tanjong Dua Belas, Sepang, Labu, Dengkil, Ampang, Cheras, Kajang, Ulu Semenyih, and Batu (Gombak) which resultant of 19 hotspot mukim. Meanwhile, the other places are insignificant, with low z-scores and high p-values.

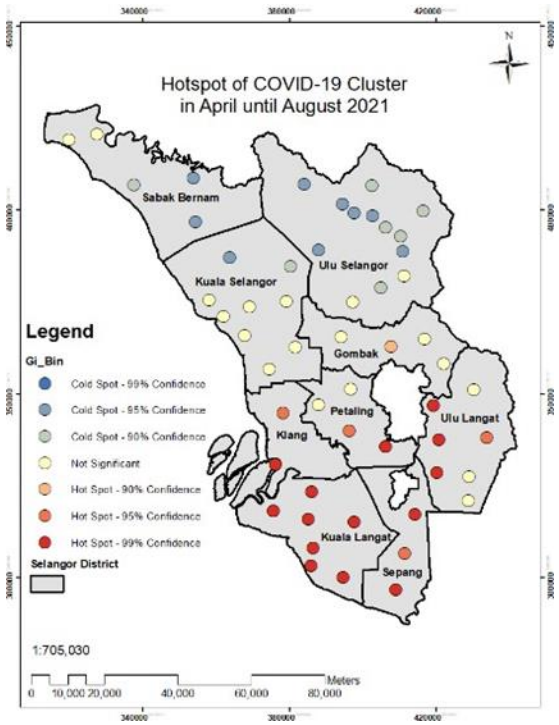


Figure 8. Hotspot detection of COVID-19 clusters from April to August 2021

Kapar, Klang, Jugra, Telok Panglima, Bandar, Kelanang, Morib, Tanjong Dua Belas, Batu (Kuala Langat), Sepang, Labu, Dengkil, Sungai Buloh, Bukit Raja, Damansara, Petaling, Kajang, Beranang, Semenyih, Cheras, Ulu Semenyih, Ulu Langat, Ampang, Ulu Kelang, Batu (Gombak), and Rawang are detected as hotspot with 99%, 95% and 90% confidence level for the COVID-19 deaths (Figure 9). Meanwhile, the cold spot area occurring in the area neighbouring boundary of Ulu Selangor, Kuala Selangor and Sabak Bernam with confidence level of 90%, 95%, and 99% which involve mukim Ulu Tinggi, Tanjong Karang, Pasir Panjang, Pancang Bedena, Sungai Panjang, Ulu Bernam, Kalumpang, Kuala Kalumpang, Sungai Gumut, and Sungai Tinggi.

Referring to Figure 5 until Figure 9, there is detected identical cold areas of COVID-19 in Sungai Besar, Pasir Panjang, Ulu Bernam, Kalumpang, Kuala Kalumpang and Sungai Gumut, while hotspot areas of COVID-19 are in Batu, Dengkil, Labu, Damansara, Petaling, Ulu Semenyih, Kajang, Cheras, Ampang, and all of the mukim in Kuala Langat and Klang.

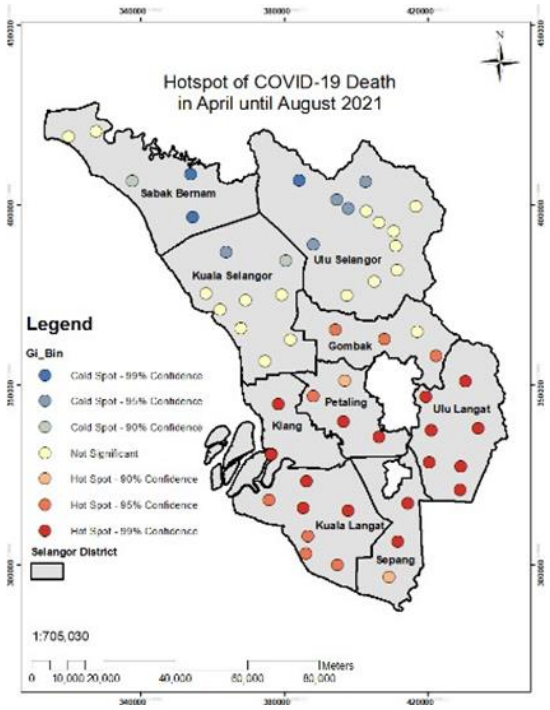


Figure 9. Hotspot detection of COVID-19 deaths from April to August 2021

4.3 Hotspot Analysis Between Parameters

The increasing of COVID-19 deaths in Malaysia, especially in Selangor, has become a concern for the people. Therefore, this study will also analyze the hotspot mapping between the parameters to identify and conclude the parameter contributing to the number of deaths. To realize the study's objective, the generated hotspot mapping determines the parameter that contributes the most to COVID-19 deaths. Based on the highly significant hotspot of COVID-19 deaths (Figure 9), it is detected that the new cases (Figure 6) and cumulative cases (Figure 7) have the same significant value in the areas which are Klang, Tanjong Dua Belas, Bandar, Telok Panglima Garang, Dengkil, Ampang, Cheras, Petaling, Damansara, and Kajang.

On the other hand, based on the COVID-19 deaths for a highly significant cold spot of 99% confidence level, new cases and cumulative cases have the exact location with the same considerable level: Ulu Bernam, Sungai Panjang and Pasir Panjang. Since new cases and cumulative cases contain the exact area of the most intense high and cold spot with COVID-19 deaths, it can be said that new cases and cumulative cases contribute the most to COVID-19 deaths. Besides that, by overlaying the pattern of the outbreak in COVID-19 deaths (Figure 9) hotspot mapping with the other parameter (Figure 5, Figure 6, Figure 7, and Figure 8), there are eight

mukim detected as high-risk areas, which are Klang, Tanjong Dua Belas, Bandar, Telok Panglima Garang, Dengkil, Ampang, Cheras and Kajang. The significant hotspot of COVID-19 death occurs due to the high value of COVID-19 cases, which are new cases and cumulative cases. On the other hand, those eight mukim tend to have significant hotspots because of the high population density. Ganasegeran et al. stated in 2022 that their study confirmed that population density directly correlated with COVID-19 infections. Hence, these eight mukim can be recognized as the areas that should receive more attention, have more impact, and are highly susceptible in the near future.

5.0 Conclusion and Recommendations

5.1 Conclusion

In this study, data processing and analysis were performed using ArcMap 10.4, the Getis–Ord Gi* and Global Moran's I model being the main tools. The main objective of this study has been achieved where the hotspot map has been generated and to analyze that new and cumulative cases are contributing to higher COVID-19 deaths. The main objective can be achieved by assisting the other two objectives, which are determining the parameters for the hotspot analysis and developing a geospatial database for the hotspot analysis. The study parameters have been identified: new cases, cumulative cases, number of death cases, number of clusters and population density, with the number of deaths as a primary parameter.

Throughout this study, it is shown that the most challenging part in this study is the data collection phase. This phase is the most difficult because there is a bunch of COVID-19 data for five months in each district that needs to be collected and processed. Hence, applying technology nowadays teaches people how to work more efficiently and quickly.

The outcome of this study, which is the hotspot analysis of COVID-19, is important because it can help to identify areas with high risk and low risk, and the government and citizens can take preventive steps in risky areas. Other than that, the current study of hotspot analysis can help to recognize which places in a study area should receive more attention, which locations are more impacted, and which areas are highly susceptible in the near future.

5.2 Recommendations

Many things can be fixed and improved in this study to make it more helpful for people. Hence, this section will provide recommendations for future studies. For this study, it is recommended that

another parameter, the population of older adults, be added to determine whether it can contribute to COVID-19 deaths. This parameter cannot be added to this study because the latest data on older adults in the Selangor district has not yet been updated, wherein the only data found was collected in 2010. On the other hand, it would be more beneficial if this study were done based on the latest scope of time or hotspot on live cases. Since this study began in 2021, it can only provide data for that year.

Acknowledgement

The authors would like to acknowledge the financial support from the Ministry of Education Malaysia under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2021/WAB04/UTM/02/2) and Universiti Teknologi Malaysia under R.J130000.7852.5F485.

Reference

Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, 27(2), 93-115.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Studies in Operational Regional Science. Springer Netherlands.

Cliff, A. D., & Ord, J. K. (1981). Spatial Processes: Models & Applications. Pion.

Columbia Public Health. (2023). Hotspot analysis. Columbia University Mailman School of Public Health. https://www.publichealth.columbia.edu/

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. Wiley.

Fatima, M., Butt, I., & Arshad, S. (2021). Geospatial clustering and hot spot detection of malaria incidence in Bahawalpur district of Pakistan. GeoJournal.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley.

Ganasegeran, K., Jamil, M. F. A., Appannan, M. R., Ch'ng, A. S. H., Looi, I., & Peariasamy, K. M. (2022). Spatial Dynamics and Multiscale Regression Modelling of Population Level Indicators for COVID-19 Spread in Malaysia. International Journal of Environmental Research and Public Health, 19(4), 2082.

Getis, A., & Ord, J. K. (1996). Local spatial statistics: An overview. In P. Longley & M. Batty (Eds.), Spatial Analysis: Modelling in a GIS Environment (pp. 261-277). GeoInformation International.

Hakim, A. (2021, January 30). COVID-19: Why Selangor has the most cases & Which district is being hit the worst. TRP. https://www.therakyatpost.com/news/malaysia/2021/01/30/covid-19-why-selangor-has-the-most-cases-and-which-district-is-being-hit-the-worst/

Li, Z., Zhang, X., & Ding, Q. (2020). Spatial Analysis of COVID-19 Clusters and Context Factors in the US. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3572341

Moran, P. A. P. (1948). The interpretation of statistical maps. Journal of the Royal Statistical Society: Series B (Methodological), 10(2), 243-251.

Parvin, F., Ali, S. A., Hashmi, S. N. I., & Ahmad, A. (2021). Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique. Spatial Information Research.

Rex, F. E., Borges, C. A. D. S., & Käfer, P. S. (2020). Spatial analysis of the COVID-19 distribution pattern in São Paulo State, Brazil. Ciencia & Saude Coletiva, 25(9), 3377-3384.

Shariati, M., Mesgari, T., Kasraee, M., & Jahangiri-rad, M. (2020). Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system (March and April, 2020). Journal of Environmental Health Science and Engineering, 18(2), 1499-1507.

Xu, H., Croot, P., & Zhang, C. (2021). Discovering hidden spatial patterns and their associations with controlling factors for potentially toxic elements in topsoil using hot spot analysis and K-means clustering analysis. Environment International, 151, 106456.

Zhou, Z., Zhang, X., Yang, Z., Chen, Y., Liu, Y., Wen, J., & Chen, W. (2020, October). Visual abstraction of geographical point data with spatial autocorrelations. In 2020 IEEE Conference on Visual Analytics Science and Technology (VAST) (pp. 60-71). IEEE.