



Faculty of Computer Science and Information Technology

A Spatial Panel Data Modeling for COVID-19 Areal Data in Sarawak

Saira Aslam

**Master of Science
2024**

A Spatial Panel Data Modeling for COVID-19 Areal Data in Sarawak

Saira Aslam

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Computational Modelling)

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK

2024

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

A handwritten signature in black ink, appearing to read 'Saira', is written over a horizontal dotted line. The signature is stylized with a star-like mark above the 'i' and a long, sweeping underline.

Signature

Name: Saira Aslam

Matric No.: 21020296

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date: 10 July 2024

ACKNOWLEDGEMENT

I am profoundly grateful to the individuals who have made significant contributions to the completion of this thesis.

First and foremost, I would like to express my deepest gratitude to Allah for His blessings, guidance, and strength throughout this research endeavor. His unwavering presence and divine support have been my constant inspiration and motivation.

I sincerely thank my supervisor, Dr. Piau Phang, and my co-supervisor, Prof. Jane Labadin, for their invaluable support, encouragement, supervision, and insightful suggestions throughout this research work. Their unwavering guidance and continuous assistance have been instrumental in enabling me to successfully complete this work. I am sincerely thankful for their patience and understanding.

I am also indebted to the Zamalah Siswazah UNIMAS Scholarship (ZSU) for providing me with the financial support that has made my studies possible. Without their assistance, I would not have been able to pursue my academic journey to this extent.

Above all, I am deeply grateful to my family, especially my Father and Mother, whose love, support, and patience have been immeasurable. This dissertation is dedicated to them as a token of my appreciation for their unwavering belief in me and their constant encouragement. Last but not least, certainly, here's a slightly expanded version of the acknowledgment statement: "I want to express my profound gratitude to all those who have played a role, whether significant or small, in shaping the outcome of this research. Your contributions have been invaluable."

ABSTRACT

The world has been facing its biggest infectious disease outbreak, the COVID-19 pandemic, since 2020. As of December 23, 2022, over 650 million confirmed cases and over 6 million deaths have been recorded worldwide. In Malaysia, with a population of 32.78 million, there have been over 5 million cases and 36,000 deaths. One of the essential properties of pandemics is their spatial spread, which depends on the disease transmission mechanism, human mobility, and control strategies. Although the disease was well controlled in Sarawak in 2020, by the end of 2021, Sarawak recorded over 250,000 infections and 1,619 fatalities, making it the second state with the highest cumulative cases among Malaysia's 13 states and three federal territories. This research quantifies the spatial autocorrelation of Sarawak COVID-19 district data across different pandemic waves to understand the spatial spread of COVID-19 in Sarawak. Regression models such as ordinary least square(OLS), spatial lag model (SLM), and spatial error model (SEM) were used to determine the spatial relationship between district-wise COVID-19 cases and various demographic, social, and economic factors. The study found that the percentage of households with garbage collection facilities, population density, and proportion of males are key factors driving COVID-19 incidence rates in Sarawak. Furthermore, this research examined spatial autocorrelation at two levels: within the state of Malaysia and the district of Sarawak. The impact of flight routes on COVID-19 transmission from state to district was analyzed using supra-adjacency matrices to compute local and global Moran's I values. Centrality measures identified pivotal nodes in districts, states, and state-to-district connections, revealing spatial dependence across these levels. Moran's I statistics indicated positive spatial autocorrelation for all levels except for states in wave 1. Significant p -values (< 0.05) for districts, two-level, and state-to-district comparisons highlighted the non-random nature of the spatial patterns, suggesting that

neighboring areas share similarities in incidence rates. Kuala Lumpur and Kuching exhibited the highest significance in state-to-district connections, Pahang at the state level, and Sibu and Sri Aman at the district level. The main contribution of this research lies in its detailed analysis of spatial patterns and factors influencing COVID-19 spread in Sarawak. The findings can assist the Government of Malaysia in devising targeted interventions and policies to control future outbreaks by understanding the spatial dynamics of disease transmission.

Keywords: COVID-19 pandemic, Sarawak, spatial autocorrelation, spatial regression model, spatial lag, spatial error model

Pemodelan Data Panel Spatial untuk Data Areal COVID-19 di Sarawak

ABSTRAK

Dunia telah menghadapi wabak penyakit berjangkit terbesarnya, pandemik penyakit koronavirus 2019 (COVID-19), sejak 2020. Sehingga 23 Disember 2022, lebih 650 juta kes yang disahkan dan lebih 6 juta kematian telah direkodkan di seluruh dunia. Di Malaysia, dengan populasi 32.78 juta, terdapat lebih 5 juta kes dan 36,000 kematian. Salah satu sifat penting pandemik ialah penyebaran spatialnya, yang bergantung pada mekanisme penghantaran penyakit, mobiliti manusia, dan strategi kawalan. Walaupun penyakit itu dikawal dengan baik di Sarawak pada 2020, menjelang akhir 2021, Sarawak merekodkan lebih 250,000 jangkitan dan 1,619 kematian, menjadikannya negeri kedua dengan kes terkumpul tertinggi di kalangan 13 negeri dan tiga wilayah persekutuan di Malaysia. Penyelidikan ini mengukur autokorelasi spatial data daerah COVID-19 Sarawak merentas gelombang pandemik yang berbeza untuk memahami penyebaran spatial COVID-19 di Sarawak. Model regresi seperti biasa least square(OLS),spatial lag model (SLM) dan spatial error model (SEM) telah digunakan untuk menentukan hubungan spatial antara kes COVID-19 mengikut daerah dan pelbagai faktor demografi, sosial dan ekonomi. Kajian itu mendapati peratusan isi rumah yang mempunyai kemudahan pengumpulan sampah, kepadatan penduduk, dan bahagian lelaki adalah faktor utama yang mendorong kadar kejadian COVID-19 di Sarawak. Tambahan pula, penyelidikan ini mengkaji autokorelasi spatial pada dua peringkat: dalam negeri Malaysia dan daerah Sarawak. Kesan laluan penerbangan pada penularan COVID-19 dari negeri ke daerah dianalisis menggunakan matriks bersebelahan untuk mengira nilai Moran's I tempatan dan global. Langkah-langkah kepusatan mengenal pasti nod penting dalam daerah, negeri dan sambungan negeri ke daerah, mendedahkan pergantungan spatial merentas peringkat ini. Statistik I Moran

menunjukkan autokorelasi spatial positif untuk semua peringkat kecuali untuk keadaan dalam gelombang 1. Nilai-p yang ketara (<0.05) untuk perbandingan daerah, dua peringkat dan negeri-ke-daerah menyerlahkan sifat bukan rawak bagi corak spatial, mencadangkan bahawa kawasan jiran berkongsi persamaan dalam kadar kejadian. Kuala Lumpur dan Kuching mempamerkan kepentingan tertinggi dalam hubungan negeri-ke-daerah, Pahang di peringkat negeri, dan Sibu dan Sri Aman di peringkat daerah. Sumbangan utama penyelidikan ini terletak pada analisis terperinci corak spatial dan faktor yang mempengaruhi penyebaran COVID-19 di Sarawak. Penemuan ini boleh membantu Kerajaan Malaysia dalam merangka intervensi dan dasar yang disasarkan untuk mengawal wabak masa depan dengan memahami dinamik spatial penularan penyakit.

Kata kunci: Pandemik COVID-19, Sarawak, autokorelasi spatial, model regresi spatial, lag spatial, model ralat spatial

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER 1 INTRODUCTION	1
1.1 Background Study	1
1.2 Research Questions	7
1.3 Objectives	7
1.4 Scope	7
1.5 Significant	8
1.6 Organization of the Thesis	8
CHAPTER 2 LITERATURE REVIEW	11
2.1 Spatial transmission of COVID-19 in the world, Malaysia and Sarawak	11
2.1.1 Socio-demographic factors	13
2.1.2 Correlation in COVID-19 cases	14
2.2 Spatial patterns of COVID-19 transmission	19

2.3	Summary	30
CHAPTER 3 METHODOLOGY		31
3.1	Data	31
3.1.1	Study Area	31
3.1.2	Data Extraction	33
3.2	Data Pre-Processing	35
3.3	Spatial autocorrelation analysis	36
3.3.1	Spatial Dependency	37
3.3.2	Contiguity Weights	37
3.3.3	Spatial Weight Matrix	38
3.3.4	Spatial Autocorrelation	39
3.3.5	Global Spatial Autocorrelation Analysis	40
3.3.6	Local Spatial Autocorrelation Analysis	40
3.4	Non-Spatial and Spatial regression models	42
3.4.1	Ordinary Least Squares (OLS)	42
3.4.2	Spatial Lag Model (SLM)	43
3.4.3	Spatial Error Model (SEM)	45
3.5	Summary	46
CHAPTER 4 SPATIAL AND NON-SPATIAL AUTO-CORRELATION ANALYSIS OF DISTRICT-WISE INCIDENCE USING REGRESSION MODELS		48

4.1	Data Analysis / Variable Description	50
4.2	Change-point Analysis and Normality Checking for Dependent Variables	53
4.3	Constructing Spatial Weights Matrix in R Programming	57
4.4	Global Spatial Autocorrelation Analysis	59
4.5	Local Indicators of Spatial Association (LISA)	62
4.6	Global Non-spatial / Spatial Regression Models (OLS, SLM, and SEM)	63
4.7	Summary	70
CHAPTER 5 SPATIAL AUTO-CORRELATION ANALYSIS ON STATE- WISE AND DISTRICT-WISE INCIDENCE USING SUPER ADJACENCY MATRIX		72
5.1	Local Moran's I for State, Two-Level and State to District	72
5.1.1	LISA for State of Malaysia	74
5.1.2	LISA for Two-level	77
5.1.3	LISA for State to District	81
5.2	Global Moran's I value	86
5.3	Centrality of Weight Matrix	89
5.3.1	Centrality of State to District (Flight Route)	91
5.3.2	Centrality of State to State	92
5.3.3	Centrality of District to District (Sarawak)	94
5.4	Summary	96
CHAPTER 6 CONCLUSION AND RECOMMENDATION		99

6.1	Contribution	100
6.2	Limitations	102
6.3	Future Works	102
	REFERENCES	104

LIST OF TABLES

	Page
Table 2.1: Summary of the articles to address problem statement	17
Table 2.2: The brief reviews of selected articles to highlight the methodology	22
Table 3.1: Description of independent variables	34
Table 4.1: Descriptive statistics of district-wise COVID-19 confirmed cases in Sarawak, categorized by East, Central, and West zones	51
Table 4.2: Summary of different waves during the COVID-19 pandemic in Sarawak throughout 2021	53
Table 4.3: Global spatial autocorrelation analysis using first-order queen contiguity spatial weights	60
Table 4.4: The coefficient estimates and statistical test values for OLS	64
Table 4.5: The coefficient estimates and statistical test values for SLM based on first-order queen contiguity	66
Table 4.6: The coefficient estimates and statistical test values for SEM based on first-order queen contiguity	68
Table 4.7: AIC and SSE results of all three models for all waves	70
Table 5.1: Incidence rate of States of Malaysia for the year and all three waves	76
Table 5.2: Incidence rate of districts in Sarawak for the year and all three waves	78
Table 5.3: Connectivity matrix of AirAsia domestic flight route	82
Table 5.4: Connectivity matrix of Malaysian Airline domestic flight route	83
Table 5.5: Connectivity matrix of state to district using both domestic flight route	85
Table 5.6: Global Moran's I values of district, state, two-level and state to district of 2021 and different wave of 2021	87

LIST OF FIGURES

	Page
Figure 1.1: The graph of cumulative cases among 13 states and 3 federal territories in Malaysia as of (a) 31 Dec 2020, (b) 31 Dec 2021.	3
Figure 3.1: Flow diagram of chapter 3	32
Figure 3.2: Study area of Sarawak's 40 districts	33
Figure 3.3: Queen and rook contiguity (Barreal et al., 2023)	37
Figure 4.1: Flow diagram of chapter 4	49
Figure 4.2: The abrupt changes in mean and variance for daily COVID-19 confirmed cases in Sarawak throughout 2021	53
Figure 4.3: The histograms for incidence rates in different waves and 2021	55
Figure 4.4: Histograms of log-transformed population size and COVID-19 cases for Wave 1, Wave 2, and Wave 3 in Sarawak districts	56
Figure 4.5: Bar plot that gives the distribution of link numbers among district	58
Figure 4.6: The schematic diagram for illustrating spatial connection between districts	59
Figure 4.7: Moran's I scatter plot	61
Figure 4.8: Local indicator of spatial association (LISA) cluster map for incidence rates in different periods	63
Figure 5.1: Flow diagram of chapter 5	73
Figure 5.2: Plot of weight matrix of State	74
Figure 5.3: LISA diagram of state of Malaysia	76
Figure 5.4: Plot of weight matrix of district in Sarawak	78
Figure 5.5: LISA diagram of two-level of Malaysia	80
Figure 5.6: LISA diagram of State to District of Malaysia	86

Figure 5.7: Centrality between states of Malaysia and the district of Sarawak according to flight route	91
Figure 5.8: Centrality between states to state of Malaysia	93
Figure 5.9: Centrality between districts to district of Sarawak state	96

CHAPTER 1

INTRODUCTION

This chapter presents an overview of modelling COVID-19 areal data in Sarawak and discusses the problem statements and objectives of the study. This chapter is divided into six sections. Section 1.1 presents the study's background, Section 1.2, and 1.3 presents the outline of the research question and objectives respectively, and Section 1.4 presents the scope of the study. Section 1.5 discusses the study's significance and an overview of the thesis organization is provided in Section 1.6.

1.1 Background Study

The pandemic of this century, coronavirus disease 2019 (COVID-19), has become a significant challenge to the nations. On December 31, 2019, the Chinese government first reported a COVID-19 outbreak in Wuhan, the capital of Hubei Province in China. The outbreak rapidly spread from Wuhan into all provinces of China and at least 24 countries (Dayun et al., 2020). Later, the virus moved to different countries through human body transmission, and international passengers worked as carriers unknowingly and unintentionally (Dutta et al., 2021). The World Health Organization (WHO) declared the epidemic as a Public Health Emergency of international concern on January 30, 2020, and a pandemic on March 12, 2020 (Sarkar et al., 2021). Since the onset of 2020, the global impact of the COVID-19 has been significant, affecting nearly every corner of the world.

Malaysia reported its first cases of COVID-19 on January 25, 2020, when two Chinese tourists from Wuhan entered the country through Singapore. From January 25 to February 16, 2020, the first wave of COVID-19 cases in Malaysia had only 22 imported and local cases. The second wave of the outbreak occurred between February 27 and the end of

August, resulting in 9,340 confirmed cases and 127 deaths in the country. The majority of these cases were linked to a religious mass gathering that took place in Sri Petaling, Selangor, between February 27 and March 3, 2020. The gathering had an estimated 14,500 local and 1,500 overseas attendees. To control the spread of the virus, Malaysia implemented a nationwide Movement Control Order (MCO) on March 18, 2020. Subsequent Conditional (CMCO) and Recovery MCOs (RMCO) proved to be successful in reducing the number of cases and deaths. On July 1, 2020, Malaysia reported zero cases of COVID-19 (Cheong et al., 2022).

As of December 23, 2022, confirmed cases of COVID-19 worldwide exceeded 650 million, with over 6 million deaths. Malaysia, with a population of 32.78 million, recorded more than 5 million cases and 36 thousand fatalities. Sarawak, the largest state in Malaysia by area, is divided into 12 administrative divisions and 40 districts, remained COVID-free until March 13, 2020, when the first case was detected. Regrettably, Sarawak recorded its first COVID-19 fatality just four days later. In response, the Government of Malaysia implemented a country-wide lockdown, known as the MCO, on the following day. Different phases of MCOs were implemented throughout the year 2020.

In Sarawak, the COVID-19 outbreak was well-controlled in 2020, with only 1,141 confirmed cases and 19 fatalities recorded. The cases were mainly concentrated in urban districts. However, at the beginning of 2021, the situation worsened, and almost all districts, including rural ones, were severely affected. At the end of 2021, Sarawak recorded more than 250,000 confirmed cases, with 1,619 fatalities, ranking second-highest among 13 states and three federal territories in Malaysia (see Figure 1.1). On April 25, 2021, the accumulated confirmed cases in the whole of Sarawak state increased tremendously, reaching 28,307, with a total of 156 fatalities (Phang et al., 2021).

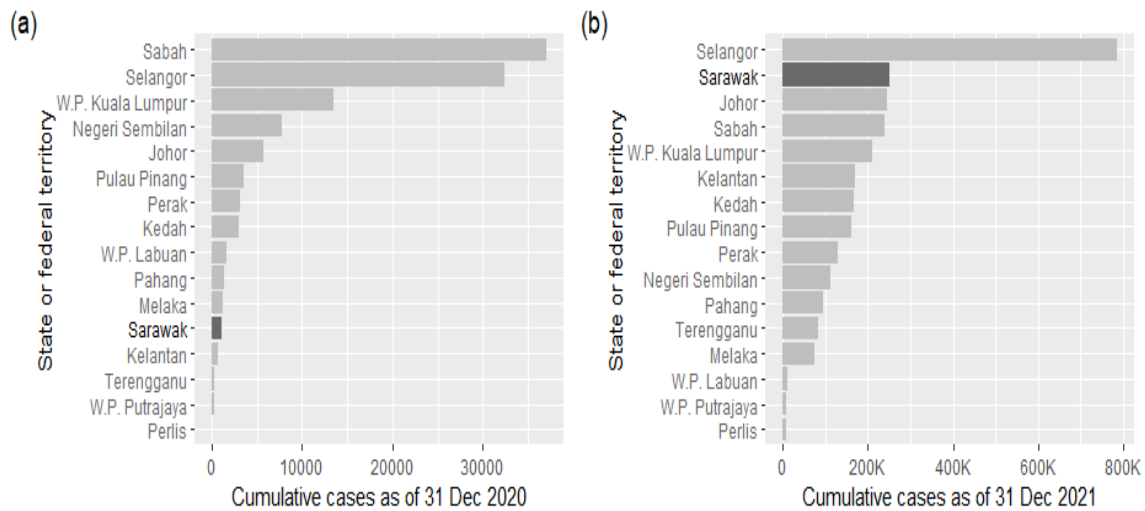


Figure 1.1: The graph of cumulative cases among 13 states and 3 federal territories in Malaysia as of (a) 31 Dec 2020, (b) 31 Dec 2021.

The spread of COVID-19 in Sarawak exhibits spatial and temporal variations at the district level. To effectively manage the pandemic, the Sarawak Disaster Management Committee (SDMC) plays a crucial role in disseminating official information, updating guidelines, and implementing preventive measures mandated by the Sarawak state government. These measures are tailored to address the evolving COVID-19 situations locally within Sarawak. The regional autonomy regulations specific to Sarawak are significant due to the significant proportion of the population residing in rural areas, where access to healthcare resources and facilities may be limited. This highlights the need for localized strategies and interventions to effectively control the spread of the virus and protect the communities in Sarawak.

The COVID-19 pandemic has taught us that societal inequality and disadvantages significantly impact how the virus spreads and its effects (Duckett et al., 2021). To understand these effects better, the study on the relationship between factors like socio-demographics and COVID-19 cases at the district level in Sarawak is vitally important.

Previous studies in Malaysia have looked at this relationship, but they only focused on the early stages of the pandemic until February 2021. Our study aims to go further by examining the entire year of 2021 and considering different factors that might influence the spread of the virus, as suggested by other researchers (Ganasegeran et al., 2022). By doing this, It is hoped that how socio-demographic factors contribute to the COVID-19 situation in Sarawak can be better understood.

Examining the spatial auto-correlation between disease rates across different regions and its relationship with socio-demographic factors over three distinct time frames are the focus of this study. Advanced statistical methods are used to analyze data and identify the most influential socio-demographic factors. By considering the spatial relationships between districts, this study can gain a better understanding of the variation in COVID-19 risk throughout Sarawak. The findings of this research will provide insight into strategies for preventing and controlling COVID-19 and other infectious diseases. Hence, this study contributes to disease surveillance and response efforts through the use of spatial analysis and geographical information.

Several recent reviews (see for instance, Franch-Pardo et al. (2020, 2021)) have emphasized the importance of a spatiotemporal analysis of disease transmission to better understand the spatial spread of the COVID-19 pandemic. Spatiotemporal data analysis is an emerging research area due to the development and application of novel computational techniques that enable the analysis of extensive spatiotemporal data. Spatiotemporal analyses are advantageous over purely spatial or time-series analyses because they allow for the investigation of both spatial and unusual patterns over time. Conducting a spatiotemporal analysis of COVID-19 cases at the district level in Sarawak is essential because it can reflect the temporal and spatial evolution of COVID-19, identify significant clusters of the disease,

explore potential variation rules, and determine whether observed space-time patterns of the epidemic are due to chance or randomly distributed. In addition, comprehensive spatiotemporal information about COVID-19 cases is essential for detecting new active and emerging clusters which can inform decision-makers and stakeholders where and when to improve targeted response measures to mitigate further transmission. Thus, a spatiotemporal investigation can prioritize districts for targeted prevention and control measures, rapid testing, and healthcare resource allocation.

Spatial data typically assumes dependence (autocorrelation) between independently measured values observed in geographical space. Therefore, a spatial autocorrelation test is necessary for any investigation of spatial dependence. Spatial autocorrelation refers to the relationship between the values of a variable (y) in location i with itself and another neighboring location j within the same geographical space of interest. Global measures of spatial autocorrelation, such as Geary's C and Moran's I statistics, including their corresponding randomization tests (permutation inference), are used to measure spatial autocorrelation. The global Moran's I value ranges from -1 to 1. A negative value of 1 means dispersion (clustering of dissimilar values), and a positive value of 1 indicates clustering (similar values). A value close to 0 or 0 implies complete spatial randomness without autocorrelation. On the other hand, local measure, namely Local Indicator of Spatial Association LISA yield multiple location-specific measures of spatial dependence, allowing researchers to assess nonstationary or how spatial dependence varies throughout space. Spatial dependence is present in many geographical datasets and exemplifies Tobler's First Law of Geography - things closer together are more similar than things further apart. Semi-variance analysis assesses scales (distances) within which spatial dependence occurs and is a method used to evaluate scales of spatial dependence.

One of the most essential properties of pandemics is their spatial spread, which mainly depends on the disease transmission mechanism, human mobility, and control strategy. Despite some publicly available descriptive statistics on district-wise confirmed cases of COVID-19 control interventions in Sarawak, no spatiotemporal pattern for COVID-19 at the district level in Sarawak is being investigated for different waves of a pandemic. Hence, this research examines the spatial dependence of the COVID-19 areal data for different administrative units in Sarawak by performing spatial panel data modeling. Specifically, the spatial autocorrelation analysis will be carried out to study the extent to which the district-wise cases are spatially correlated. Then, spatial regression models are constructed to validate the crucial role of spatial proximity in the diffusion of the pandemic at the district level in Sarawak. It is hoped that the spatial panel data modeling may assist the local public health authorities in imposing applicable disease control intervention at the district level and preparing for the next pandemic.

Since one of the most essential properties of pandemic is their spatial spread, which mainly depends on the disease transmission mechanism, human mobility and control measures, the COVID-19 situation at district level in Sarawak cannot be fully comprehended without considering the COVID-19 intensity at other states of Malaysia. Even though the whole Malaysia was under partially lockdown situation (i.e. MCO 2.0) especially in the first half of 2021, travelling across states via flight connections for essential activities was still possible. Such human mobility plays a very important role in spillovering the diseases from states with flight routes to certain Sarawak's districts, especially districts with flight connections to other states namely Kuching, Sibu, Bintulu and Miri. Hence, it becomes important to examine how this human mobility can be conceptualized by introducing a

supra-adjacency type of spatial weight to quantify the spatial dependence of states level in Malaysia to district level in Sarawak.

1.2 Research Questions

This study aims at answering the following three research questions (RQ).

RQ1: How did the evolution of COVID-19 cases in Sarawak unfold across different waves of the pandemic?

RQ2: What is the extent of spatial dependence of district-wise COVID-19 confirmed cases in Sarawak and state-wise in Malaysia, as quantified through spatial panel data modeling?

RQ3: How do demographic, social, and economic factors spatially relate to district-wise COVID-19 cases in Sarawak, as determined by spatial regression models?

1.3 Objectives

The research objectives (RO) of this study are given below.

RO1: To describe the evolution of Sarawak COVID-19 data in different waves of pandemic.

RO2: To quantify the spatial dependence of district-wise in Sarawak and state-wise in Malaysia aggregated COVID-19 confirmed cases using spatial panel data modelling.

RO3: To construct spatial regression models for determining the spatial relationship between district-wise COVID-19 cases and demographic, social, and economic factors.

1.4 Scope

The study focuses on confirmed COVID-19 cases across all 40 districts in Sarawak from January 1 to December 31, 2021. Global and local spatial autocorrelation analysis are conducted to examine the strength of spatial dependence of COVID-19 confirmed cases at district level in Sarawak. Also, how such spatial dependence can be altered if the COVID-

19 situation at state level in Malaysia is also examined through the use of supra-adjacency matrix. Additionally, various district-level demographic data (such as age structures and race) and socioeconomic indicators (like average household income) will be utilized to construct a spatial regression model. The COVID-19 data are collected from daily press statements published by the SDMC and GitHub of Ministry of Health Malaysia, while the Census data are from the Department of Statistics Malaysia. The data analysis and modeling carried out using R-studio by utilizing spdep, splm, etc. packages.

1.5 Significant

The findings of this study will be helpful in better understanding the spatial dependency of disease transmission at district and state-level, as well as the relationship between pandemic, demographic, and socio-economic factors at the district level in Sarawak. This insight is essential for decision-making in healthcare sectors, including clinical settings, medical services, medicines, public health, and other associated segmental resources among districts in Sarawak.

1.6 Organization of the Thesis

This thesis comprises six chapters, for which the summary of the next five chapters are discussed below.

CHAPTER 2 This study reviews the existing literature on COVID-19 case forecasting worldwide, with a specific focus on predicting the spread of the virus in Sarawak, Malaysia. The review identifies key research gaps and outlines a suitable methodological approach to model the spatiotemporal dynamics of COVID-19 cases in the region.

The review examines the common data preprocessing techniques employed in COVID-19 forecasting studies, such as calculating incidence rates and constructing spatial weight matrices. It also evaluates the application of spatial autocorrelation analysis and

various regression modeling techniques, including spatial and non-spatial models, to predict COVID-19 cases.

By synthesizing the insights from the literature, the study establishes a comprehensive framework for COVID-19 case prediction in Sarawak. This includes data preprocessing steps, spatial autocorrelation analysis, and the selection of appropriate regression models to capture the underlying spatial and sociodemographic factors influencing the transmission of the virus.

CHAPTER 3 presents the methodology to propose a prediction system for COVID-19 cases. The steps involved in the methodology are pre-processing spatial auto-correlation and regression. The data is divided into three periods/waves using change point analysis, incidence rate, and normality log-transformation in the pre-processing category: first-order spatial contiguity weights, Global Moran's I, LISA cluster map under spatial autocorrelation. Furthermore, SLM, SEM regression methods under spatial, and OLS under non-spatial are used.

CHAPTER 4 mentioned that the analysis was conducted using spatial autocorrelation analysis and regression models, which include the global non-spatial linear, global spatial lag, and spatial error.

CHAPTER 5 mentioned that the focus was on examining the global and local Moran's I value for Malaysia, considering the two-level and state-to-district scenarios. Before computing the global and local Moran's I value, the weight matrix for both state and two-level analyses was thoroughly examined. The supra adjacency method also introduced the weight matrix for the connections from the state to the district level. Lastly, the degree, betweenness, and eigenvector centrality were assessed for flight routes from the state to the district, state to state, and district to district.

CHAPTER 6 concludes the research by recapitulating the objectives of Chapter 1 and discussing how each was achieved. This chapter also outlines the study's contribution and limitations and suggests other aspects of the subject domain that require future research.

CHAPTER 2

LITERATURE REVIEW

This chapter comprehensively evaluates global and Malaysian studies on the spatial dynamics of COVID-19 transmission. It emphasizes the impact of the pandemic worldwide and underscores the significance of spatial analysis techniques in understanding disease spread and its correlation with demographic and socioeconomic factors. The reviewed studies highlight the importance of population density, urbanization, and spatial proximity in shaping COVID-19 transmission patterns. This chapter's insights will inform evidence-based decision-making and targeted interventions to control the spread of the virus.

2.1 Spatial transmission of COVID-19 in the world, Malaysia and Sarawak

The world has been facing numerous infectious diseases spreading, ranging from influenza, SARS, MERS, Zika, and the recent COVID-19 pandemic (Chakraborty et al., 2020; Lupton, 2021). The pandemic has caused high and far-reaching social, economic, and demographical impacts on nearly all the countries in the world. COVID-19 pandemic, caused by SARS-CoV2, is challenging global health systems since it was first reported from Wuhan, China (Goscé et al., 2020; AlQadi et al., 2021). The pandemic has significantly affected people's mobility global social, environmental, and economic (Ahmar et al., 2020; Bilal et al., 2020), leading to changes in daily life and routine behaviours.

The first COVID-19 case in Malaysia was reported on January 25, 2020, involving three Chinese tourists who entered the country via Johor from Singapore on January 23. Since then, Malaysia has experienced three major waves of COVID-19 outbreaks. The first wave occurred between January 25 and February 16, 2020, before the WHO officially classified COVID-19 as a pandemic. The second wave occurred between February 27 and

June 30, 2020, while the third wave began on September 8, 2020 and 31st December 2020 (Zamri et al., 2021). Rampal and his team reported a decline in the number of new COVID-19 cases daily in May 2020, suggesting that the peak was over but not the end of the outbreak (Rampal et al., 2021).

The third wave, which began on September 8, 2020, is the most severe of the three, with consistent daily records of more than 1000 new cases since November 3, 2020 (Dong et al., 2022). Although the COVID-19 mortality rate in Malaysia is comparably lower than in many other countries, with only 1272 deaths recorded out of 345,000 confirmed cases as of March 31, 2021, or 0.37% of detected cases (Aw et al., 2021), the pandemic is still straining the country's health systems and economy. According to the official statistical data from the MOH, Malaysia, the first wave recorded 22 infected individuals and 0 deaths. The second wave recorded cumulative cases of 8,616 infected individuals and 121 deaths, while the third wave recorded a total of 103,551 infected individuals and 343 deaths as of December 31, 2021.

Sarawak, the largest state in Malaysia, was also affected by the COVID-19 pandemic, with its first COVID-19 confirmed cases reported on March 13, 2020, and Malaysia's first COVID-attributable death on March 17, 2020 (Yang et al., 2020). Sarawak is divided into 12 administrative divisions and 40 districts, each with different population sizes, medical resources, and accessibility.

COVID-19 could be considered well controlled in Sarawak in 2020. However, at the end of year 2021, Sarawak recorded more than 250,000 cases with 1,619 fatalities, which ranked Sarawak the second-highest cumulative cases among 13 states and three federal territories in Malaysia. The management of the COVID-19 pandemic in Sarawak state is overseen by the SDMC. The SDMC plays a crucial role in disseminating official news,

providing updates, and implementing the Standard Operating Procedures (SOPs) for preventive measures mandated by the Sarawak state government. It is worth noting that some of these SOPs are specifically designed to address the evolving COVID-19 situations at the local level within Sarawak. These district-specific regulations hold great importance for Sarawak, given that 45.3% of its population resides in rural areas (DOSM, 2022), often lacking adequate medical resources and facilities.

2.1.1 Socio-demographic factors

As COVID-19 is an infectious disease that spreads from person to person, they are very natural evolve in space and time, whereby different spatial distribution patterns can be observed over time (Ben Abacha et al., 2021). Specifically, spatiotemporal analysis becomes the critical element to better understanding the spread of COVID-19 at various spatial resolution levels and for different waves of COVID-19 (Muus et al., 2021). The COVID-19 pandemic has highlighted the significant impact of inequality and structural disadvantage on its progression and outcomes (Duckett et al., 2021). Exploring indicators of structural disadvantage during an emerging pandemic is crucial, and an ecological study can effectively examine these factors. In Sarawak, where COVID-19 and census data are available at the district level, it is practical to conduct an ecological study investigating the relationship between socio-demographic factors and COVID-19 cases across all 40 districts. While previous studies in Malaysia have examined spatial modeling and regression at the district level (Zakaria et al., 2021; Cheong et al., 2022; Ganasegeran et al., 2022), their analysis only covered the disease's evolution until February 2021. Therefore, this present study builds upon previous research (Phang et al., 2021) in Sarawak by considering four distinct temporal periods throughout 2021 and incorporating various socio-demographic indicators suggested by Ganasegeran et al. (2022).

Several studies have been conducted to predict COVID-19 cases in different countries, by considering their unique demographics, population density, and other factors that influence the spread of the virus (Wong et al., 2023). For example, Sarkar et al. (2021) investigated and identified demographic risk factors associated with COVID-19 transmission in 64 districts of Bangladesh using spatial and statistical modeling. Their findings revealed significant clustering of population density based on spatial autocorrelation analysis conducted. Similarly, Gupta et al. (2022) explored the impact of socio-economic factors, such as job insecurity, overcrowding, and homelessness the spread of COVID-19 cases.

2.1.2 Correlation in COVID-19 cases

In Malaysia, studies have also been conducted to predict COVID-19 cases nationally. Ganasegeran et al. (2022) used SAR and GWR to predict COVID-19 cases, although their analysis was based on a dataset spanning only 14 days (January 22, 2021, to February 4, 2021). Jayaraj et al. (2021) employed Bayesian methods to predict COVID-19 cases over a longer duration (from March 16, 2020, to May 31, 2021). Faziera et al. (2020) utilized a Geographical Information System (GIS) approach to analyze COVID-19 cases in all 145 Malaysia districts. Zakaria et al. (2021) conducted Exploratory Spatial Data Analysis (ESDA) and applied Moran's I statistic to 144 districts in Malaysia. Similarly, Aw et al. (2021) and Iderus et al. (2022) employed Pearson's coefficient to predict COVID-19 cases based on two dataset from 2020 and covering September 20, 2020, to September 19, 2021, respectively.

Cheong et al. (2022) utilized Moran's I and the LISA to predict COVID-19 cases in 36 districts, including Peninsular Malaysia to the west and Sabah and Sarawak to the east of the South China Sea. The study duration spanned from January 22 to February 24, 2021.

Another study focused explicitly on Sarawak, employing spatial autocorrelation analysis and a spatial panel regression model to predict COVID-19 cases using a 16-week dataset from January to April 2021 (Phang et al., 2021). It is evident that no study focused explicitly on Sarawak for a longer period of time. Spatial Autocorrelation, a concept related to geographical clustering, has been applied to studying the spread of infections among neighboring districts with similar demographic or social features, including in the case of the Coronavirus (Alkhamis et al., 2020; Kang et al., 2020). Recent research has investigated the impact of population, city size, demographics, and transportation systems on the spread and mortality rate of diseases such as plague (Reyes et al., 2013), smallpox (Reyes et al., 2013), influenza (Reyes et al., 2013), SARS (Reyes et al., 2013), COVID-19 (Li et al., 2020; Ribeiro et al., 2020; Xie et al., 2020). To combat infectious diseases like COVID-19, it is crucial to understand the dynamics of spatial spread and implement appropriate preventive measures (Li et al., 2020; Zhou et al., 2020). Extending the concept of spatial Autocorrelation to determining the mean center of wealth and infections has enabled the analysis of the temporal dynamics of wealth distribution, economic growth, and infectious diseases (Grether et al., 2010). Recent publications have explored other aspects of the spread of COVID-19 in Lebanon, including preventive measures, psychological and pharmaceutical responses, and mental health needs (El Deeb, 2021). Therefore, the proposed study aims to predict COVID-19 cases in Sarawak using data from 2021. Details of all these articles are presented in Tables 2.1 and 2.2 for emphasizing different aspects: Table 2.1 outlines the problem statement, demonstrating the importance of analysing the spread of the virus in Sarawak, while Table 2.2 discusses the techniques that can be employed for this analysis. These tables collectively provide a comprehensive overview of the rationale behind

the study and the methodologies applicable for examining the COVID-19 data at district level in Sarawak.

Table 2.1: Summary of the articles to address problem statement

No	Main Author / Year of publication / Title of Article	Dependent (D)/ independent variable (I)	Study area; its spatial resolution	Methods	Study period; data source	Main finding/conclusion
1	Sarkar et al. (2021)	D: Cases I: 9 Demographic factors	Bangladesh: 64 districts	Spatial autocorrelation, Moran's I statistics; spatial regression	2020; Institute of Epidemiology Disease Control and Research (IEDCR)	- Population density played a significant role in explaining COVID-19 transmission patterns in Bangladesh
2	Gupta et al. (2022)	D: cases I: Proportion of urban population, roads per hundred sq. km, airport availability etc.	India: 640 districts	LISA spatial autocorrelation technique	22 Mar -11 Aug 2020;	- The crucial role of spatial proximity in pandemic diffusion was validated. - Higher urbanization rates were positively correlated with a higher incidence of infections, as revealed by spatial regression models.
3	Ganasegeran et al. (2022)	D: cases I: regression indicators	Malaysia; 14 days	SAR, geographically weighted regression (GWR),	22 Jan - 4 Feb 2021; Ministry of Health Malaysia	- COVID-19 was concentrated mostly in Malaysia's Central and Southern regions.
4	Jayaraj et al. (2021)	D: cases, mortalities, testing I: Population	Malaysia	Bayesian	16 Mar 2020 - 31 May 2021;	- The epidemiology of COVID-19 is constantly evolving and shows no signs of slowing down.
5	Faziera et al. (2020)	D: cumulative number of COVID-19 positive cases I: Population and Housing Census	Malaysia: 145 districts	GIS	25 Jan to 12 May of 2020	- The spatial distribution of COVID-19 is expected to provide valuable information for implementing preventive measures among Malaysia's population.
6	Iderus et al. (2022)	D: cumulative number of COVID-19 cases, I: population and population density	Malaysia: 145 districts	Pearson's correlation	20 Sep 2020 - 19 Sep 2021;	- More populous and densely populated districts have a higher risk of COVID-19 transmission, especially with the Delta variant becoming the dominant circulating strain.

Table 2.1 continued

7	Zakaria et al. (2021)	D: cases I:	Malaysia; 144 districts	ESDA, Moran's I statistic	17 Apr -31st Oct 2020;	- COVID-19 hotspots have undergone significant changes over time. - Positive spatial autocorrelation suggests a non-random distribution of COVID-19 occurrences.
8	Aw et al. (2021)	D:cumulative COVID-19 cases, infection rates of COVID-19 cases I:Population Density	Malaysia; 143 districts	Pearson's coefficient	19 Jan to 31 Dec 2020	- Passive containment measures such as urban planning strategies provide valuable non-pharmaceutical lessons for disease interventions against future contagious diseases.
9	Cheong et al. (2022)	D: cases I: deaths	Malaysia; 36 districts	Moran's I and LISA	22 Jan - 24 Feb 2021;	- The spatiotemporal transmission risk pattern is useful in monitoring the pandemic's spread.
10	Phang et al. (2021)	D: cases I:recovered, deaths	Sarawak; 16 weeks	Spatial autocorrelation analysis, spatial panel regression model	Jan - Apr 2021;	- COVID-19 spreading in neighboring districts significantly affects the rate of confirmed cases in specific districts of Sarawak.

Based on the Table 2.1, Sarkar et al. (2021) conducted a study in Bangladesh, analyzing COVID-19 transmission patterns across 64 districts using spatial autocorrelation and Moran's I statistics, with a focus on demographic factors. Their findings indicated that population density significantly influenced COVID-19 transmission patterns in Bangladesh during 2020. Gupta et al. (2022) examined the pandemic diffusion in India, covering 640 districts. Using LISA spatial autocorrelation techniques, they analyzed the influence of urban population, road density, and airport availability on COVID-19 cases from March to August 2020. Their study validated the crucial role of spatial proximity in the spread of the pandemic and found that higher urbanization rates were positively correlated with a higher incidence of infections.

Ganasegeran et al. (2022) explored COVID-19 cases in Malaysia over a 14-day period in early 2021. Using SAR and geographically weighted regression (GWR) methods, they concluded that COVID-19 cases were concentrated mostly in Malaysia's Central and Southern regions. Jayaraj et al. (2021) utilized Bayesian methods to study COVID-19 cases, mortalities, and testing in Malaysia from March 2020 to May 2021. They found that the epidemiology of COVID-19 was constantly evolving and showed no signs of slowing down.

Faziera et al. (2020) conducted a study in Malaysia, focusing on 145 districts. Using GIS methods, they examined the cumulative number of COVID-19 positive cases from January to May 2020. Their findings suggested that the spatial distribution of COVID-19 provided valuable information for implementing preventive measures among Malaysia's population. Iderus et al. (2022) investigated the cumulative number of COVID-19 cases in 145 districts of Malaysia from September 2020 to September 2021. Using Pearson's correlation, they concluded that more populous and densely populated districts had a higher risk of COVID-19 transmission, especially with the Delta variant becoming dominant.

Zakaria et al. (2021) analyzed COVID-19 cases across 144 districts in Malaysia using ESDA and Moran's I statistics. Their study period spanned from April to October 2020, and their findings indicated significant changes in COVID-19 hotspots over time, with positive spatial autocorrelation suggesting a non-random distribution of COVID-19 occurrences.

Aw et al. (2021) focused on cumulative COVID-19 cases and infection rates in 143 districts of Malaysia from January to December 2020. Using Pearson's coefficient, they highlighted the importance of urban planning strategies as passive containment measures, providing valuable non-pharmaceutical lessons for future contagious diseases. Cheong et al. (2022) studied COVID-19 cases and deaths in 36 districts of Malaysia using Moran's I and LISA methods from January to February 2021. Their findings emphasized the usefulness of spatiotemporal transmission risk patterns in monitoring the pandemic's spread.

Phang et al. (2021) conducted a study on COVID-19 cases in Sarawak over a 16-week period from January to April 2021. Using spatial autocorrelation analysis and spatial panel regression models, they found that COVID-19 spreading in neighboring districts significantly affected the rate of confirmed cases in specific districts of Sarawak.

In the summary of literature analysis, researches were focused on the spatial analysis of COVID-19 transmission in different countries, particularly Bangladesh, India, and Malaysia. The review of Table 2.1 highlights a significant gap in studies focused on Sarawak over a longer period. Previous research has predominantly covered broader regions or shorter time frames. For instance, studies by Sarkar et al. (2021) and Gupta et al. (2022) focused on COVID-19 transmission patterns in Bangladesh and India, respectively, utilizing spatial autocorrelation techniques. Studies in Malaysia by Ganasegeran et al. (2022), Jayaraj et al. (2021), and others primarily analyzed COVID-19 cases across various districts but did not extend their focus specifically to Sarawak for an extended duration.

Phang et al. (2021) did study COVID-19 cases in Sarawak, but their analysis spanned only a 16-week period in early 2021. Given this context, the proposed study aims to predict COVID-19 cases in Sarawak using data from 2021, providing a comprehensive and extended analysis that fills the existing research gap.

2.2 Spatial patterns of COVID-19 transmission

Spatial patterns of COVID-19 transmission provide critical insights into how the virus spreads geographically, influenced by a complex interplay of socio-demographic, environmental, and behavioral factors. The spatial distribution of cases can vary significantly across regions, reflecting population density, urbanization levels, healthcare infrastructure, and adherence to public health measures. Understanding these patterns is crucial for devising targeted interventions and allocating resources effectively to mitigate transmission risks and support healthcare systems.

Table 2.2 presents a comprehensive overview of studies that utilize spatial analysis techniques to unravel the dynamics of COVID-19 transmission. These studies employ diverse methodologies such as spatial autocorrelation, spatial regression models, and geographical information systems (GIS) to map out spatial clusters and identify hotspots of infection. For instance, studies like Kang et al. (2020) and Gulyev (2020) use spatial autocorrelation to detect clusters and examine how demographic factors like population density contribute to localized outbreaks. Similarly, Dutta et al. (2021) and Al-Kindi et al. (2020) employ spatial lag models and GIS to explore the influence of urbanization and population density on virus spread within specific geographic contexts.

Table 2.2: The brief reviews of selected articles to highlight the methodology

No	Main Author / Year of publication / Title of Article	Dependent/ independent variable	Study area; its spatial resolution	Methods	Tools / Software	Study period; data source	Main finding/conclusion
1	Kang et al. (2020)	D: Cases I: Population, population density, No of doctors, No of hospital beds	China: 31 provinces	Spatial epidemic dynamics, Moran's I spatial statistic	Statistica 1 software R	16 Jan - 6 Feb 2020;	- It is highly probable that COVID-19 can spread between regions that are near each other.
2	Dutta et al. (2021)	D: Cases I: Deaths	India: 640 districts	Spatial lag and error models; Geographical weighted regression (GWR); LISA	Geoda 1.12; GWR 4.0 / MGWR 2.2	Up to 3 Jun 2020; https://www.covid19india.org	- COVID-19 cases are most susceptible to population density, urbanization, and access to banking facilities.
3	Guliyev (2020)	D: Cases I: Death, Recovered	China: 31 regions	6 spatial panel data models, i.e. Spatial Autoregression / error / Durbin models etc.	Stata 16	22 Jan - 10 Mar 2020;	- The number of COVID-19 cases increases with the rate of deaths while the number of recovered cases decreases.
4	Adekunle et al. (2020)	D: cases deaths, I: no of countries,	Africa: 52 states	Dynamic panel data system generalized method of the moment (GMM)	Statistics		- A linear correlation exists between confirmed cases of COVID-19 and attributable deaths.
5	Alexakis et al. (2021)	D: lockdown I: direct and indirect (spillover) effects	Measures on 45 stock market	Spatial Durbin Model (DSDM) with fixed effects Spatial Durbin Model (DSDM) with fixed effects		2 Jan - 8 Apr 2020;	- There exists a negative relationship between stock market returns and the intensity of lockdown measures.
6	Mitze et al. (2022)	D: I:	Germany: 3 regions	Spatial panel error correction model		https:// doi. org/ 10. 6084/ m9. figsh are. 12789 941.	- Local hotspots play a crucial role in the spread of the disease, and containment measures effectively mitigate the spread.

Table 2.2 continued

7	Al-Kindi et al. (2020)	D: cases I: population size, population density,	Oman: 61 wilayats	Geographical Information System (GIS) techniques	ArcPro 2.5	29 Apr - 30 Jun 2020;	- Different districts have varying rates of COVID-19 infections, with significant spatial variability.
8	Osayomi et al. (2021)	D: Vaccine I: 16 (Socio-environmental)		Global Moran's I, Anselin, local Moran, Pearson correlation and multivariate regression.		Africa Centre for Disease Control (Africa CDC),	
9	Boumahdi et al. (2021)	D: I: infected rate, hospitalization rate, mortality rate	Morocco: 75 provinces	Moran's spatial autocorrelation,	R-Studio	2 Mar - 12 May 2020;	- The epidemic spreads hierarchically between large cities and through transport routes. - The diffusion mode through districts has contributed to the emergence of family outbreaks despite strict containment measures.
10	Ma et al. (2021)	D: Cases I:	China: 362 cities, 31 provinces	Statistical charts, hotspot analysis, spatial autocorrelation, and Poisson space-time scan statistic.	ArcGIS Desktop software (version 10.3.1,	10 Jan - 5 Oct 2020;	- Scientific references can help in allocating medical resources and monitoring potential rebounds of the COVID-19 epidemic in China.
11	Troppy et al. (2021)	D: I: unemployed population > 16 years, population aged > 65 (total 11 indicators)	Massachusetts : 351 towns	GWR, Moran's I spatial autocorrelation	R version 4.04	1 Feb - 17 June 2020;	- Social vulnerability indicators are associated with variations in SARS-CoV-2 testing rates.
12	Wu et al. (2021)	D: fatality I: Demographical factors	Texas: 254 counties	OLS, GWR, Spatial autocorrelation,			- Geographically Weighted Regression (GWR) models provide more geodata-oriented information than Ordinary Least Squares (OLS) models.
13	Sun et al. (2020)	D: cumulative confirmed cases I: sociodemographic	U. S: county-level data set for the Contiguous United States	SRM, Moran's I statistic		28 Jun 2020;	- The model fit varies spatially, identifying regions where the fit is heterogeneous and requires closer attention in the short term.

Table 2.2 continued

14	Liu et al. (2020)	D: COVID-19 confirmed cases, I: epidemic data	Wuhan 13 districts	Spatial autocorrelation, Global Moran's I,		29 Jan - 29 Feb 2020;	- COVID-19's evolution exhibits spatial concentration and temporal continuity.
15	Huang (2021)	D: I:	Internet (4000 sampling locations around the world)	the heat map, global Moran's I, geographic mean center,	Python, ArcGIS	22 Jan 2020 - 22 Jan 2021;	- The spatial autocorrelation of confirmed cases increased gradually from the initial outbreak until September 2020 and decreased slightly.
16	Wong et al. (2020)	D: confirmed cases I: population density		diagnostic tests		Feb - May 2020;	
17	Briz-Redón et al. (2022)	D: I: socio-demographical data	Valencia, Spain, 14 municipal districts		R studio	19 Feb - 31 Aug 2020;	- Classical and non-sophisticated methods like the first-order contiguity or inverse distance matrix are still appropriate.
18	Yi et al. (2021)	D: I:	Singapore		ArcGIS	22 to 26 Apr 2020;	
19	Kumpel et al. (2022)	D: I:	India: Hubballi, Dharwad	Genetic matching algorithm	R studio		- It is essential to ensure adequate, uninterrupted, and reliable water supplies during public health emergencies and provide alternative supplies and sufficient household storage for households unprepared for outages.
20	Rahman et al. (2021)	D: confirmed cases, deaths, recovery I:	Bangladesh: 64 districts			8 Mar - 8 Aug 2020;	
21	Tammes (2020)	D: COVID-19 cases I:	England			16 Mar -19 Apr 2020;	

Table 2.2 continued

22	Coşkun et al. (2021)	D: I: demographic and climate	Turkey: 81 provinces	Regression analysis		Mar 2020;	- The finding that COVID-19 can spread invisibly in the air and more in windy weather suggests that the virus in the atmosphere is a significant threat to humans, particularly with increasing air circulation due to wind speed.
23	Kan et al. (2021)	D: confirmed cases I:	Hong Kong: 18 districts			23 Jan, 14 Apr, 6 Jul- Aug 29 of 2020;	- Policymakers should focus on targeting mitigation measures in high-risk areas and vulnerable groups and provide guidance to the public to avoid visiting and conducting activities in high-risk places.

Table 2.2 presents a comprehensive review of various studies analyzing the spread of COVID-19 using different statistical and spatial methodologies. Each study's choice of approach reflects the unique requirements of the study area, and data availability. This critical analysis evaluates the rationale behind the selected approaches and their effectiveness in understanding COVID-19 dynamics.

Kang et al. (2020) used spatial epidemic dynamics and Moran's I spatial statistic to study COVID-19 cases in China. The use of Moran's I is particularly appropriate for identifying spatial autocorrelation, which helps determine whether COVID-19 cases are randomly distributed or clustered in specific areas. This method is effective in regions with varying population densities, like China's provinces, to highlight clusters of infection.

Dutta et al. (2021) employed spatial lag and error models, GWR, and LISA for analyzing COVID-19 cases across India's 640 districts. The spatial lag and error models account for spatial dependencies and heterogeneity, while GWR provides localized insights, revealing how relationships between variables change across space. LISA maps visualize local clusters of high or low incidence, making these methods suitable for a large and diverse country like India, where spatial dynamics can significantly differ.

Guliyev (2020) utilized spatial panel data models, including Spatial Autoregression and Durbin models, to study COVID-19 cases in China. These models capture the spatial and temporal dimensions of the data, addressing both the spatial spillover effects and time-lagged influences. This approach is valuable for understanding how past infections influence current cases across regions, offering a dynamic perspective on the epidemic's evolution.

Adekunle et al. (2020) applied the dynamic panel data system GMM to analyze cases and deaths across 52 African states. GMM is effective for handling potential endogeneity issues and unobserved heterogeneity in panel data. It provides robust estimates by

accounting for the correlation between the lagged dependent variable and the error term, which is crucial for understanding the dynamics of COVID-19 spread in a continent with diverse healthcare systems and reporting standards.

Alexakis et al. (2021) employed the Spatial Durbin Model to study the effects of lockdown measures on stock markets. The Spatial Durbin Model captures both direct and indirect (spillover) effects, making it suitable for understanding how lockdown measures in one region can affect stock markets in another. This approach highlights the interconnectedness of economic activities and public health measures, which is particularly relevant in a globalized economy.

Mitze et al. (2022) used spatial panel error correction models to analyze local hotspots in Germany. This method is adept at modeling long-term equilibrium relationships and short-term dynamics, providing insights into how local outbreaks influence the broader epidemic trends. It is effective for identifying persistent hotspots and assessing the impact of containment measures over time.

Al-Kindi et al. (2020) used GIS techniques to study spatial variability in COVID-19 infections in Oman. GIS provides powerful tools for visualizing and analyzing spatial data, making it suitable for identifying geographic patterns and hotspot areas. This approach is valuable for public health planning and targeted interventions in regions with significant spatial heterogeneity.

Osayomi et al. (2021) employed Global Moran's I and multivariate regression to analyze vaccination rates in Africa. Moran's I identifies spatial clustering of vaccination rates, while multivariate regression models assess the influence of socio-environmental factors. This combination provides a comprehensive view of the spatial distribution and determinants of vaccination coverage.

Boumahdi et al. (2021) and Ma et al. (2021) applied spatial autocorrelation techniques to study COVID-19 spread in Morocco and China, respectively. These methods are effective for detecting non-random spatial patterns and understanding the spread mechanisms, such as hierarchical diffusion and transport routes, which are critical for designing effective containment strategies.

Troppy et al. (2021) used GWR and Moran's I spatial autocorrelation to analyze social vulnerability indicators in Massachusetts. GWR offers localized insights, revealing how social vulnerability varies across towns, while Moran's I identifies clustering patterns. This dual approach helps in tailoring public health interventions to the specific needs of different communities.

Wu et al. (2021) compared OLS and GWR models for analyzing fatality rates in Texas. OLS provides a global perspective, while GWR offers localized insights, highlighting the spatial variability in the relationship between demographical factors and fatality rates. This comparison underscores the importance of choosing appropriate models based on the spatial characteristics of the data.

In conclusion, various studies such as Kang et al. (2020) and Guliyev (2020) utilized spatial autocorrelation and panel data models to identify spatial clustering and temporal dynamics, highlighting the influence of population density and other demographic factors on COVID-19 transmission. Similarly, Dutta et al. (2021) and Al-Kindi et al. (2020) employed spatial lag models and GIS techniques to explore the impact of urbanization and population density on virus spread. The discussion emphasized the significant role of socio-demographic factors in the spread of COVID-19 in Sarawak.

To understand COVID-19 patterns in Sarawak, three core components were examined: the waves of COVID-19, spatial autocorrelation analysis, and spatial regression

models. The review suggests using change point analysis to identify the COVID-19 waves in 2021, which would help in understanding temporal shifts in infection rates and the impact of different waves on the overall pandemic dynamics.

Furthermore, studies like those by Dutta et al. (2021) and Boumahdi et al. (2021) effectively detected clusters and spatial patterns, providing insights into how COVID-19 cases are distributed across different regions. Consequently, spatial autocorrelation analysis methods such as first-order spatial contiguity weights, Global Moran's I, and LISA maps are recommended.

Researchers like Guliyev (2020) and Ma et al. (2021) accounted for spatial dependencies and heterogeneity, offering a nuanced understanding of the factors driving COVID-19 spread in different areas. For regression analysis, the Ordinary Least Squares (OLS) method is advised for non-spatial analyses as it provides a straightforward approach to understanding the relationship between COVID-19 cases and socio-demographic factors. However, for spatial regression models, the Spatial Lag Model (SLM) and Spatial Error Model (SEM) are recommended.

In summary, the analysis of Table 2.2 underscored the significant role of socio-demographic factors in the spread of COVID-19 in Sarawak. To understand these patterns, three core components were examined: the waves of COVID-19, spatial autocorrelation analysis, and spatial regression models. The review suggests using change point analysis to identify the COVID-19 waves in 2021. Additionally, it recommends first-order spatial contiguity weights, Global Moran's I, and LISA maps for spatial autocorrelation analysis. For regression analysis, the OLS method is advised for non-spatial analyses, while the SLM and SEM are recommended for spatial regression models.

2.3 Summary

This chapter explored efforts to understand the spread of COVID-19 globally, with a specific focus on Malaysia and particularly Sarawak. The discussion highlighted the significant role of socio-demographic factors in the spread of COVID-19 in Sarawak. To comprehend the patterns of COVID-19 in Sarawak, three core components were examined: the waves of COVID-19, spatial autocorrelation analysis, and spatial regression models. The review suggests using change point analysis to identify the COVID-19 waves in 2021. Additionally, it recommends using first-order spatial contiguity weights, Global Moran's I, and LISA maps for spatial autocorrelation analysis. For regression analysis, the OLS method is advised for non-spatial analyses, while the SLM and SEM are recommended for spatial regression models.

CHAPTER 3

METHODOLOGY

This chapter is divided into five parts: data collection, pre-processing, spatial autocorrelation, and regression models. It explains the assembly of datasets, division of the study period into three segments, and measurement of spatial autocorrelation. The chapter also quantifies the spatial dependence of COVID-19 cases at the district and state levels using spatial panel data modeling techniques, exploring the relationships between dependent and independent variables. Additionally, both spatial and non-spatial regression models are constructed to investigate the impact of demographic, social, and economic factors on district-wide COVID-19 cases in Sarawak. Fourteen demographic and socioeconomic variables were selected for this analysis. The flow diagram of this chapter is presented in Figure 3.1.

3.1 Data

In the data collection process data extraction on the daily press statements released by the SDMC is described. Subsequently, data pre-processing is performed and COVID-19 incidence rates are calculated.

3.1.1 Study Area

Sarawak is the largest state in Malaysia by land area and is separated from Peninsular Malaysia by the South China Sea. Sarawak is divided into 12 administrative divisions and further subdivided into 40 districts (see Figure 3.2), in which Kuching city is the capital of Sarawak. Sarawak shares an interstate border with Sabah via the Lawas district, national borders with Brunei via districts Miri, Marudi, Limbang, and Lawas in the north, and Indonesia via the Sarawak-Kalimantan border. As of 2020, the population of Sarawak was

estimated to be around 2.45 million, which ranked Sarawak the fifth highest population by state in Malaysia. Even so, Sarawak has the lowest population density, with 20 persons per square kilometer (DOSM, 2021), and nearly 57.5% of the Sarawak population resides in five urban districts: Kuching, Samarahan, Sibuan, Bintulu, and Miri.

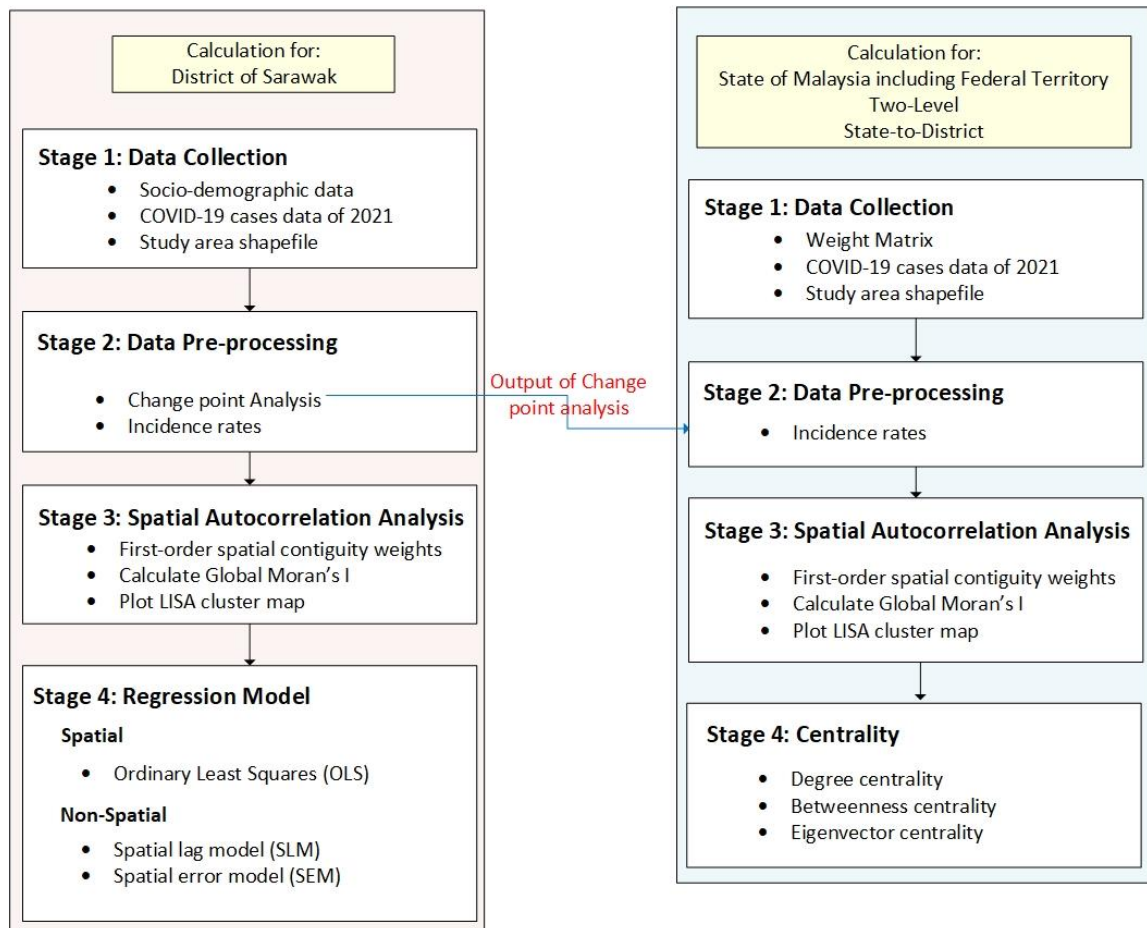


Figure 3.1: Flow diagram of chapter 3

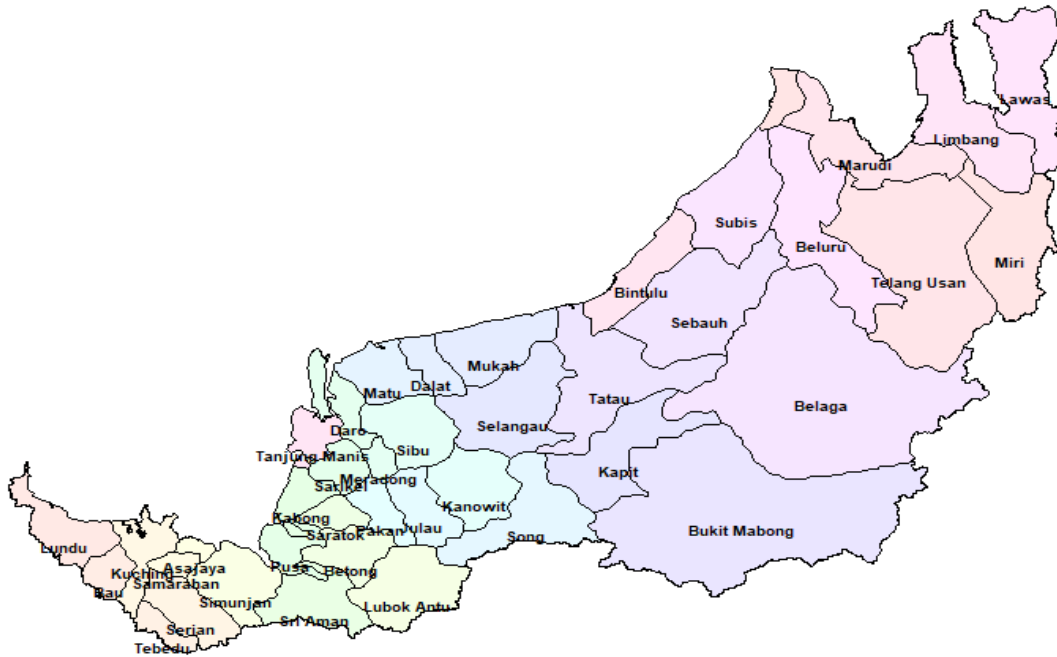


Figure 3.2: Study area of Sarawak's 40 districts

3.1.2 Data Extraction

District-wise COVID-19 data of Sarawak were extracted from the daily press statements published by the SDMC, and the data for new cases at state level of Malaysia and total population has been sourced from the official website of the MOH and Department of Statistics Malaysia (DOSM), respectively. The daily COVID-19 confirmed cases from 1 January to 31 December 2021 are collected for all 40 districts in Sarawak and all states including three federal territories. The socio-demographic covariate data are obtained from the DOSM. The administrative boundary shapefile of Sarawak districts is obtained from an open license resource known as Geo Boundaries Global Administrative Database (Runfola et al., 2020). Table 3.1 lists all fourteen independent variables; the household income and basic amenities data per administrative district in Sarawak are derived from the survey report released by the DOSM, while the Population and Housing Census of Malaysia 2020 provided the demographic data (DOSM, 2022).

Table 3.1: Description of independent variables

Category	Variable name	Abbreviation	Variable description
Demographic	Population growth rate	Pop_GrowthR	The rate of increase or decrease in the number of residents compared to the previous year.
	Non-citizen population	Pop_Noncitizen	The percentage of residents who are not citizens.
	Male population	Pop_Male	The percentage of male residents.
	Population between 15 and 64	Pop_15to64	The percentage of residents between the ages of 15 and 64.
	Population 65 years and above	Pop_65	The percentage of residents who are 65 years old or above.
	Population density	Pop_Density	The average number of residents per square kilometer.
	Household size	Household_size	The average number of people living in a dwelling unit.
(Household) Income-related criteria	Median of monthly income	Median_inc	The median monthly household income.
	Gini coefficient	Gini	A measure that represents income inequality.
	Incidence of poverty	Poverty	The proportion of households with income below the poverty line.
Basic amenities/urban infrastructure	Internet subscription	Internet	The percentage of households with an internet subscription at home.
	Piped water	Piped_water	The percentage of households with piped water supply in the house.
	Public health centers	Public_h	The percentage of households located within 5 km of the nearest public health center.
	Garbage collection facility	Garbage	The percentage of households with garbage collection facilities within 100 meters of their living quarters.

The variables include population growth rate, non-citizen population, male population, population between 15 and 64, population 65 years and above, population density, household size, median monthly income, Gini coefficient, incidence of poverty, internet subscription, piped water, public health centres, and garbage collection facilities.

These variables were selected because they reflect the socioeconomic and demographic characteristics of the population, which are known to impact the transmission dynamics of the disease. For instance, population density and garbage collection facilities are important indicators of urbanization and infrastructure quality, which can affect the spread of the disease. Similarly, household size, median monthly income, and Gini coefficient are indicators of socioeconomic status, which can influence the ability of individuals to access healthcare and follow public health guidelines. By incorporating these variables into the analysis, the study aimed to provide a comprehensive understanding of the sociodemographic factors that contribute to the spread of COVID-19 in Sarawak, Malaysia.

Due to data availability, this study did not consider environmental factors, transport, and mobility elements at the district level. Due to the different orders of magnitude for independent variables, the standardization procedure will need to carry out before modeling. While the list may not be exhaustive, all variables (except the first four and Median_inc) are indicators of social vulnerability in existing COVID-19 research (Fallah-Aliabadi et al., 2022). The first four variables and Median_inc are chosen since they are found to be significant variables in COVID-19 incidence rates in existing research (Sun et al., 2020; Zhu et al., 2020; Al Awaidy et al., 2021; Wang et al., 2022).

3.2 Data Pre-Processing

This part explores the temporal patterns of COVID-19 incidence rates across different periods, peaks, or waves in Sarawak, Malaysia. To achieve this, this study segmented the daily confirmed cases data for 2021 into three temporal periods using change-point analysis, following the methods described by Gargoum et al. (2021) and Phang et al. (2021). Following the change point analysis of Sarawak's COVID-19 data, comparable dates were selected to partition the COVID-19 data for the state of Malaysia into the same three

distinct temporal periods. This segmentation enables subsequent spatiotemporal analysis to understand the variations in incidence rates.

To ensure comparability among different districts and state with varying socio-demographic characteristics, this study apply a standardization procedure to account for the differences in orders of magnitude in the data.

District-wise and state-wise COVID-19 incidence rates are calculated by dividing the cumulative confirmed cases in a given wave by the population size of the corresponding district and state for each temporal period using Equation 3.1:

$$\text{Incidence rate} = \left(\frac{\text{the number of confirmed cases}}{\text{the population of a district}} \right) \times 1000 \quad \text{Equation 3.1}$$

Incidence rates can be used to plot histograms, which will help assess the normality of COVID-19 cases over different time periods.

If the normality assumption is unmet, a logarithm transformation is applied to the data. This comprehensive data pre-processing ensures a robust analysis of COVID-19 data and provides valuable insights into the temporal and spatial patterns of the disease's impact on different districts and populations. It is essential to clarify that change-point analysis, standardization, and incidence rate calculation procedures are distinct and sequentially carried out to achieve accurate results for subsequent analyses.

3.3 Spatial autocorrelation analysis

Spatial autocorrelation is a statistical technique used to model relationships between variables observed at different locations in space (geographical location). The tendency of nearby locations to exhibit similar values for a given variable, whereas-spatial heterogeneity, refers to the variation in spatial patterns across the study area.

3.3.1 Spatial Dependency

Spatial dependence refers to the degree of spatial autocorrelation between independently measured values observed in geographical space. Several statistics are used, including **Moran's Statistics Indices**, to assess the presence of spatial dependence or variation in a variable of interest between spatial units. However, before discussing Moran's formula, it is essential to briefly understand the concept of contiguity in geographical location and how to quantify it through a spatial weight matrix.

3.3.2 Contiguity Weights

Contiguity means that two spatial units share a common border of non-zero length. There are two common types of contiguity, namely Rook and Queen Contiguity, which can be distinguished in analogy to the moves allowed for the such-named pieces on a chess board. The rook criterion defines neighbors by having a shared edge between two spatial units. That is, neighbors in the Rook criterion have at least two common boundary points (a segment), which matches the movement of the Rook in chess. The queen criterion defines neighbors as spatial units sharing a standard edge or a common vertex that matches the movement of the Queen in chess. These two contiguities for the case of a regular grid of points are shown in Figure 3.3.

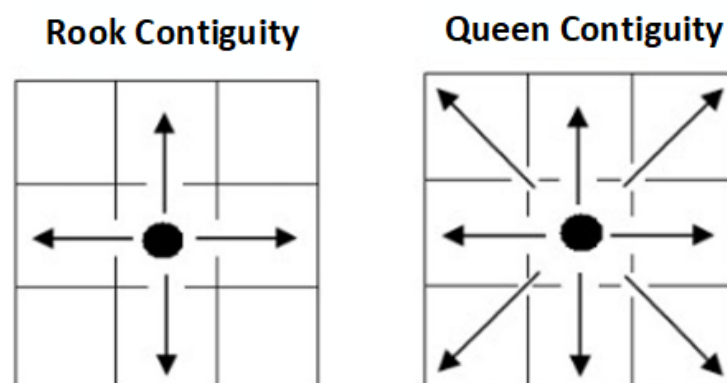


Figure 3.3: Queen and rook contiguity (Barreal et al., 2023)

3.3.3 Spatial Weight Matrix

Spatial weights are vital in any analysis of spatial dependence. They are essential in constructing spatial autocorrelation statistics and provide the means to create spatially explicit variables, such as spatially lagged variables and spatially smoothed rates. The districts' spatial association can be manifested by constructing spatial weights based on spatial contiguity (or geographical distance which is not chosen in this study). first-order queen contiguity weights will be used in both spatial autocorrelation analysis and global spatial regression model.

Formally, the weights express the neighbor structure between the observations as an $n \times n$ matrix W in which the elements w_{ij} of the matrix are the spatial weights:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \quad \text{Equation 3.2}$$

In Equation 3.2 the spatial weights w_{ij} are non-zero when i and j are neighbour, and zero otherwise. By convention, the self-neighbour relation is excluded, so that the diagonal elements of W are zero, $w_{ij} = 0$.

In its simplest form, the spatial weights matrix expresses the existence of a neighbour relation as a binary relationship, with weights 1 and 0. Formally, each spatial unit is represented in the matrix by a row i , and the potential neighbors by the column j , with $j \neq i$.

Usually, the spatial weight matrix is expressed in row-standardized form. Row-standardization takes the given weights w_{ij} (e.g, the binary zero-one weights) and divides them by the row sum, as given in Equation 3.3.

$$w_{ij}(s) = \frac{w_{ij}}{\sum_j w_{ij}} \quad \text{Equation 3.3}$$

As a result, each row sum of the row-standardized weights equals one. Also, the sum of all weights $S_0 = \sum_i \sum_j w_{ij}$, equals n , the total number of observations.

The first-order queen contiguity spatial weight is defined in Equation 3.4:

$$w'_{ij} = \begin{cases} 1, & \text{if district } j \text{ shares a common boundary with district } i \\ 0, & \text{otherwise} \end{cases} \quad \text{Equation 3.4}$$

where w_{ij} is the normalized weight in which its magnitude indicates the intensity of spatial proximity between two districts and for Sarawak $n = 40$ in this study. Assumes that the district-wise incidence rates are influenced by the first-order neighbors For instance, the Bau district has two first-order neighbors, Kuching and Lundu.

Queen contiguity spatial weights are constructed with R package `spdep` (Bivand et al., 2009). For first-order contiguity, the average number of neighbors is 4.4. Sibü is the most connected district, with eight neighbors. As no standard scientific criterion exists for selecting an optimal spatial weight matrix or neighborhood structure (Hannum et al., 2019), this study will present a spatial autocorrelation analysis for district-wise incidence rates in different temporal durations obtained from first-order spatial contiguity weights.

3.3.4 Spatial Autocorrelation

Spatial autocorrelation can be used to explore the spatial association or spatial variation over time. Specifically, for investigating the spatial autocorrelation of COVID-19, global spatial autocorrelation can be applied for daily new confirmed cases at different time points. Moran's I statistic is chosen as the index of spatial autocorrelation, and it has a value range of $[-1, 1]$. If the value is 0, there is no correlation between cases in different spatial districts (such as districts, cities, etc.), and the variable of interest, such as disease incidence, is categorized as spatially random or independent. Positive Moran's I value represents the incidence of disease in adjacent areas is similar, which means a positive spatial correlation. The closer to 1, the stronger the spatial aggregation. Moran's I value less than 0 indicates a

negative spatial correlation. The smaller the value, the more significant the spatial difference among spatial units.

3.3.5 Global Spatial Autocorrelation Analysis

The spatial autocorrelation analysis is then performed to determine to what extent the systematic spatial variation in district-wise incidence rates is presented. The following global Moran's I statistics are calculated to investigate whether the nearby districts tend to have similar incidence rates. Global Moran's I measures the overall clustering of the spatial data. It is defined as in equation 3.5.

$$I = \frac{N}{W} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad \text{Equation 3.5}$$

Where N denotes the number of spatial units, and i, j represents spatial unit indexes. The variable of interest for spatial unit i (and j) is denoted as x_i (and x_j). Here, $\bar{x}$ represents the overall variable of interest, while signifies the mean of the variable x . The matrix of spatial weights, w_{ij} , is crucial in spatial analysis, characterized by zeroes on the diagonal ($w_{ii} = 0$), facilitating the assessment of spatial relationships between different units. Additionally, the term plays a fundamental role, indicating the sum of all spatial weights within the spatial units, as denoted by $\sum_{i=1}^N \sum_{j=1}^N w_{ij}$.

3.3.6 Local Spatial Autocorrelation Analysis

Local spatial autocorrelation mainly analyzes spatial association patterns at different spatial locations to find spatial heterogeneity among data. The LISA cluster map is plotted to illustrate the type of LISA for each district with an assessment of the significant levels of local spatial statistics. Two broad categories of local spatial association exist, namely spatial cluster (positive local spatial autocorrelation) and spatial outlier (negative local spatial autocorrelation). This study used LISA to analyze the spatial cluster of incidence rate. LISA values can be interpreted as follows:

- **High-high clusters:** These are areas with a high COVID-19 incidence rate, surrounded by other areas with a high incidence rate. These areas are likely to be hotspots for COVID-19 transmission.
- **Low-low clusters:** These are areas with a low COVID-19 incidence rate, surrounded by other areas with a low incidence rate. These areas may be less affected by COVID-19.
- **High-low outliers:** These areas have a high COVID-19 incidence rate, surrounded by areas with a low incidence rate. These areas have a high possibility of spill over diseases to their neighboring areas.
- **Low-high outliers:** These areas have a low COVID-19 incidence rate, surrounded by areas with a high incidence rate. These areas may be at risk of increased COVID-19 transmission.

The disease spillover risk is higher in the districts identified as spatial outliers. Local Moran's I, also known as LISA, indicates the significance of spatial differences.

$$LISA = \frac{(x_i - \bar{x})}{S^2_x} \sum_{j=1}^n [w_{ij}(X_j - \bar{X})] \quad \text{Equation 3.6}$$

In Equation 3.6, $S^2_x = \sum (X_i - \bar{X})^2 / n$ introduces the concept of LISA. LISA are used to detect locations of spatial non-stationarity or outliers. Simultaneously, this can describe significant correlations at a specific location as local spatial clusters or correlations between observations and neighboring observations (hot spots).

Local Moran's I value is required to draw LISA for district and state-level incidence rates. Steps to calculate the Moran's I are given below:

- a) Calculate the spatial weight matrix. A common spatial weight matrix is the queen contiguity matrix, which assigns a weight 1 to each observation that is adjacent to the i^{th} observation.
- b) Calculate the mean incidence rate ($\bar{x}$).

- c) Calculate the spatial lag of the incidence rate ($w'x_i$). The spatial lag of a variable is the weighted average of the values of the variable for neighboring observations.
- d) Calculate the Moran's I value using the following formula:

$$\text{Moran's I} = \frac{(\sum w_i(x_i - \bar{x})(w'x_i - \bar{x}))}{ns^2} \quad \text{Equation 3.7}$$

where, w_i , x_i is the spatial weight, incidence rate for the i^{th} observation, respectively. $\bar{x}$, n , and s^2 represent the mean incidence rate, number of observations, and variance of the incidence rate, respectively.

3.4 Non-Spatial and Spatial regression models

Spatial regression models across multiple spatial scales have been used extensively in COVID-19 research (Guliyev, 2020). researchers commonly used global spatial regression models to examine the connection between dependent variables (e.g. incidence rates at the district level) and independent variables while considering the spatial relationship between spatial units. Spatial dependence refers to the influence that neighboring spatial units have on each other. Several different non-spatial and spatial models can be used to understand the relationship between variables and COVID-19 rates. These include OLS, SLM, and SEM, which are global models (Dutta et al., 2021). However, the non-spatial (i.e. OLS) model does not allow for investigating spillover effects as it implicitly assumes that outcomes for different spatial units are independent.

3.4.1 Ordinary Least Squares (OLS)

OLS within the framework of global regression is a fundamental and widely used statistical technique employed to analyze the relationship between a dependent variable and one or more independent variables. It aims to find the coefficients that best fit a linear equation to the observed data by minimizing the sum of the squared differences between the observed and predicted values.

In this study, OLS is used to examine the linear relationship between the district-wise incidence rates (dependent variable) and the various independent variables. By estimating the coefficients β , the OLS method provides insights into the strength and direction of the associations between the variables. It serves as a fundamental starting point in understanding the essential relationships within the data before incorporating more sophisticated techniques, such as spatial regression models, to account for spatial dependencies or errors. Mathematically, the OLS can be expressed as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \epsilon \quad \text{Equation 3.8}$$

where Y is the dependent variable, $x_1, x_2, x_3 \dots x_n$ represents the independent variables, β_0 is the intercept term, $\beta_1, \beta_2, \beta_3 \dots \beta_n$ are the coefficients corresponding to each independent variable.

OLS is a non-spatial technique that serves as the foundation for understanding primary relationships in data. However, in cases where spatial dependencies or errors are present, such as when neighboring observations influence data points, spatial regression models are necessary. In this study, two standard spatial regression techniques, namely the spatial lag model (SLM) and spatial error model (SEM) are constructed. These models account for the spatial interdependencies among spatial units (i.e. districts) and provide a more accurate representation of the underlying relationships in the data by incorporating spatial components.

3.4.2 Spatial Lag Model (SLM)

It extends the traditional OLS model by incorporating a spatial weight matrix that captures spatial relationships between observations. Estimation of coefficients considers the spatial dependence structure, enhancing the modeling accuracy. This model provides

valuable insights into spatial patterns, relationships, and the impact of neighboring observations on the dependent variable.

In Equation 3.9, the dependent variable is denoted as Y , while X represents the matrix of independent variables. The vector of regression coefficients for the independent variables is denoted as β . Introducing spatial dependence into the model, the spatial autoregressive coefficient, denoted as ρ , captures the extent of spatial correlation. The spatial lag term, W_y , is a crucial component, representing the weighted sum of the dependent variable Y for neighboring observations, based on the defined spatial weight matrix W . Finally, the error term, ϵ , is assumed to be independently and identically distributed, contributing to the overall model's residual variation.

$$Y = X\beta + \rho W_y + \epsilon \quad \text{Equation 3.9}$$

In SLM spatial dependence is explicitly considered by including a spatially lagged dependent variable on the right-hand side of the regression Equation 3.9. This approach allows us to capture the influence of neighboring districts on the dependent variable, considering spatial relationships among the observations.

$$Y = X\beta + \lambda W\rho + \epsilon \quad \text{Equation 3.10}$$

In Equation 3.10, y represents a $1 \times n$ vector of the dependent variable, I is the $n \times k$ matrix of independent variables, and W is the $n \times n$ spatial weight matrix capturing spatial relationships between districts. The coefficients β estimate the impact of independent variables, while the spatial autoregressive parameter ρ indicates how neighboring districts influence each other's incidence rates. Scalar coefficient λ that represents the spatial autoregressive parameter it measures the strength and direction of spatial dependence. The random error term ϵ accounts for unexplained variability in the model (Maiti et al., 2021).

Employing these global spatial regression models provides crucial insights into spatial patterns and relationships in the data.

The SLM considers the influence of neighboring observations on the dependent variable. It includes a spatial lag term, the weighted sum of neighboring observations, to capture the spatial spillover effects. SLM recognizes that the values can influence the values of a variable in one location in nearby locations.

By utilizing the SLM and SEM, this study gains a comprehensive understanding of the spatial relationships and patterns of COVID-19 incidence rates in the study area, considering both the direct influence of neighboring districts and unobserved spatial factors that may affect the dependent variable.

3.4.3 Spatial Error Model (SEM)

The SEM incorporates an error term that captures unobserved spatial factors influencing the dependent variable, acknowledging the correlation between error terms of neighboring observations due to spatially structured residuals. That is, the SEM account for potential spatial autocorrelation in the residuals. The SEM is represented by equation 3.11:

$$y = X\beta + u \quad \text{where } u = \lambda Wu + \epsilon \quad \text{Equation 3.11}$$

here, λ represents the autoregressive factor, and λWu represents the spatial error term. If λ is statistically significant, it indicates the presence of hidden independent variables with spatial autocorrelation, which may contribute to non-negligible spatial autocorrelation in the residuals (Ulimwengu et al., 2021). The SEM helps in exploring and understanding the spatial patterns of the unexplained variation in the data, providing valuable insights into potential spatial dependencies in the model.

3.5 Summary

This chapter presented the methodology to understand the spread of COVID-19 in Sarawak. To comprehend the patterns of COVID-19 in Sarawak, this chapter is divided into four parts: data, data pre-processing, spatial autocorrelation analysis, and non-spatial and spatial regression models. First parts explain the assembly of the datasets for this study which include the study area which is 40 district of Sarawak, and data extraction which is done from extracted from the daily press statements published by the SDMC, and the data for new cases and total population has been sourced from the official website of the MOH. In the second part of this methodology, data pre-processing is done.

The second section examines the temporal patterns of COVID-19 incidence rates across various periods, peaks, or waves in Sarawak, Malaysia. To achieve this, the study segments the daily confirmed case data for 2021 into three temporal periods using change-point analysis. Additionally, district-wise and state-wise COVID-19 incidence rates are calculated by dividing the cumulative confirmed cases in each wave by the population size of the respective district and state.

In the third part of the methodology, spatial autocorrelation is employed to model relationships between variables observed at different geographical locations. This technique accounts for spatial autocorrelation, which is the tendency for nearby locations to have similar values, and spatial heterogeneity, which refers to the variation in spatial patterns across the study area.

The final part of the methodology involves constructing both spatial and non-spatial regression models to investigate the relationships between district-wide COVID-19 cases and the districts' demographic, social, and economic factors. The study uses spatial panel regression with 14 selected demographic and socioeconomic variables to analyze their

impact on COVID-19 cases in all 40 districts of Sarawak. To understand the relationship between these variables and COVID-19 rates, the study considers OLS for non-spatial models and the SLM and SEM for spatial models.

CHAPTER 4

SPATIAL AND NON-SPATIAL AUTO-CORRELATION ANALYSIS OF DISTRICT-WISE INCIDENCE USING REGRESSION MODELS

This chapter examines the interaction between COVID-19 rates and socio-demographic factors at the district level in Sarawak. Methods include data description, change-point analysis, checking for normality with histograms, constructing spatial weight matrices, scrutinizing global spatial autocorrelation, and evaluating LISA. Regression modeling is conducted using OLS for non-spatial analyses, and SLM and SEM for spatial regression models. This framework aims to understand the interplay between COVID-19 incidence rates and socio-demographic variables from both spatial and non-spatial perspectives. The flow diagram of this chapter is presented in Figure 4.1.

The process in Figure 4.1 involves four stages: data collection, data pre-processing, spatial autocorrelation analysis, and regression modeling. In the first stage, data on socio-demographic variables and COVID-19 cases are collected. In the second stage, the COVID-19 data are divided into three periods or waves using change point analysis to identify significant changes. The data is then standardized, and normality is checked. Log-transformation is performed if the data is not normally distributed. Incident rates are calculated as the number of confirmed cases per population, scaled by 1000. In the third stage, first-order spatial contiguity weights are calculated using the Queen Contiguity method, and global Moran's I is calculated to test for spatial autocorrelation. A LISA cluster map is also plotted to visualize the spatial patterns. In the fourth stage, three regression models are used: the SLM, SEM, and OLS. The SLM accounts for spatial autocorrelation in the dependent variable, while the SEM accounts for spatial autocorrelation in the error term. The OLS model is used as a baseline model.

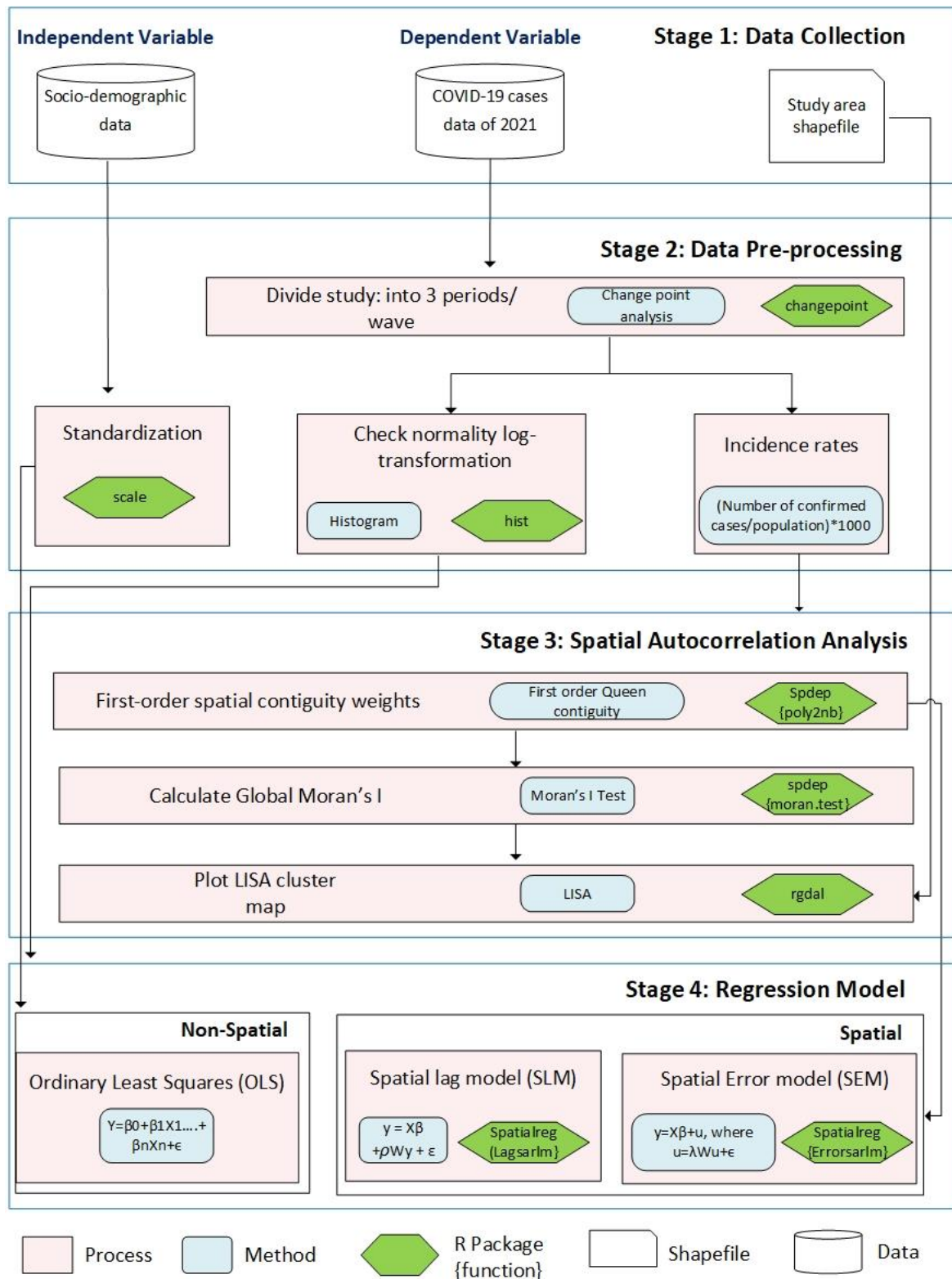


Figure 4.1: Flow diagram of chapter 4

4.1 Data Description

Table 4.1 presents the descriptive statistics of district-wise COVID-19 incidence rates throughout 2021 in Sarawak for whereby the 40 districts are divided into three zones: the East, Central, and West. These statistics provide a comprehensive overview of COVID-19 intensities within each zone. The analysis aims to shed light on the distribution and characteristics of these factors in the context of the different geographical districts.

Table 4.1 details the minimum, 25th percentile (q25), median (q50), 75th percentile (q75), IQR, maximum, mean, standard deviation, skewness, kurtosis, and mode for COVID-19 cases in each district within the respective zones. These statistics provide valuable insights into the range, central tendency, and variability of the variables under consideration.

The East zone encompasses districts such as Lundu, Kuching, Samarahan, and others. The Central zone districts, including Betong, Sibu, Kanowit, and others, exhibit distinct statistics contributing to the broader analysis. Lastly, the West zone districts, such as Bintulu, Miri, Limbang, and others, present their descriptive statistics contributing to understanding the regional differences in the variables of interest.

Notably, among the districts, Mukah has the highest COVID-19 incidence rates, indicating the relative magnitude of cases in that area. Additionally, Meradong stands out with the highest mean value, suggesting a unique characteristic in terms of health outcomes. These observations offer a glimpse into the districts' diversity of situations regarding COVID-19 and socio-demographic attributes.

Table 4.1: Descriptive statistics of district-wise COVID-19 confirmed cases in Sarawak, categorized by East, Central, and West zones

No.	East Zone	Min	q25	q50	q75	IQR	Max	Mean	Standard Deviation	Skewness	Kurtosis	Mode
1	Lundu	0	0	1	6	6	217	10.605	28.357	4.097	18.359	0
2	Bau	0	0	4	13	13	238	20.532	42.007	2.839	7.986	0
3	Kuching	0	17	65	205	188	1866	203.967	335.847	2.626	7.135	2
4	Samarahan	0	1	7	21	20	417	27.838	56.662	3.356	12.954	0
5	Asajaya	0	0	0	4	4	133	7.515	19.675	4.069	18.517	0
6	Simunjan	0	0	0	6	6	227	8.523	21.659	4.934	35.16	0
7	Tebedu	0	0	0	3	3	94	5.326	13.321	3.663	14.906	0
8	Serian	0	1	9	28	27	630	36.962	80.423	4.044	20.234	0
9	Sri Aman	0	0	4	16	16	198	17.277	33.099	2.97	9.133	0
10	Lubok Antu	0	0	0	2	2	105	3.904	10.971	4.63	27.894	0

Central Zone

11	Pusa	0	0	0	3	3	142	6.805	17.462	3.857	17.681	0
12	Kabong	0	0	0	1	1	34	1.433	3.631	4.711	29.465	0
13	Betong	0	0	2	10	10	214	12.438	27.901	3.924	18.562	0
14	Saratok	0	0	1	7	7	131	8.422	18.853	3.647	15.15	0
15	Sarikei	0	0	3	12	12	168	11.293	20.693	3.356	14.299	0
16	Pakan	0	0	0	4	4	106	4.981	12.585	4.191	20.909	0
17	Meradong	0	0	2	8	8	121	7.11	12.748	3.838	22.452	0
18	Julau	0	0	1	3	3	74	3.227	7.896	5.454	36.722	0
19	Tanjung Manis	0	0	0	0	0	68	1.57	5.474	7.5	74.917	0
20	Daro	0	0	0	0	0	61	1.145	4.339	8.542	102.91	0
21	Matu	0	0	0	1	1	27	1.597	3.781	3.724	16.536	0

Table 4.1 continued

22	Dalat	0	0	1	4	4	32	3.214	5.776	2.565	6.775	0
23	Mukah	0	0	3	18	18	175	13.408	23.356	3.263	13.953	0
24	Sibu	0	27	54	98	71	436	76.471	79.445	2.19	5.373	4
25	Kanowit	0	0	2	9	9	84	8.052	13.268	2.569	7.433	0
26	Selangau	0	0	2	7	7	111	6.482	12.771	4.601	29.294	0
27	Song	0	0	1	7	7	112	6.986	14.321	3.35	13.854	0
28	Kapit	0	1	6	21	20	156	14.858	20.168	2.585	9.937	0
29	Bukit Mabong	0	0	0	3	3	199	3.416	12.615	11.251	162.102	0
30	Belaga	0	0	0	2	2	285	5.899	20.618	8.317	96.835	0

West Zone

31	Tatau	0	0	1	6	6	93	6.474	13.514	3.258	11.904	0
32	Bintulu	0	10	28	61	51	435	50.926	67.309	2.791	9.295	1
33	Sebauh	0	0	1	2	2	57	2.43	6.03	5.674	40.676	0
34	Subis	0	0	4	11	11	160	11.515	20.884	3.722	18.557	0
35	Miri	0	10	30	82	72	361	55.537	64.497	1.844	3.658	4
36	Beluru	0	0	1	4	4	74	3.43	7.335	5.029	34.306	0
37	Marudi	0	0	0	1	1	63	1.493	4.621	8.086	92.381	0
38	Telang Usan	0	0	0	1	1	70	1.879	6.681	7.096	60.778	0
39	Limbang	0	0	0	5	5	150	7.907	17.86	3.585	17.546	0
40	Lawas	0	0	0	4	4	70	5.258	11.108	3.102	11.157	0

4.2 Change-point Analysis and Normality Checking for Dependent Variables

The COVID-19 daily confirmed cases data for 2021 are divided into three periods by performing change-point analysis. This statistical method, implemented through the changepoint package in R (Killick et al., 2014), identifies significant changes in the time series data, such as shifts in means or variances. The abrupt changes in mean and variance occurred on 5 September and 10 November (see Figure 4.2). The subsequent analysis is carried out for the three designated periods within the year, whereby their brief overview is given in Table 4.2.

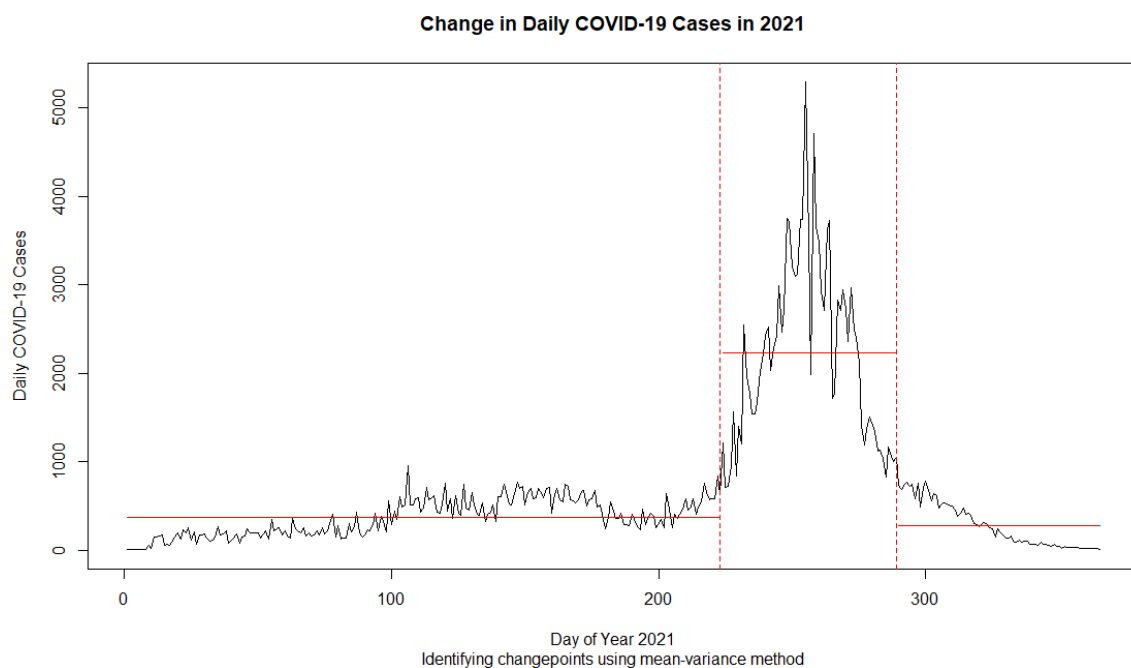


Figure 4.2: The abrupt changes in mean and variance for daily COVID-19 confirmed cases in Sarawak throughout 2021

Table 4.2: Summary of different waves during the COVID-19 pandemic in Sarawak throughout 2021

	Period	Number of days	Average number of daily cases	Range of district-wise Incidence rates
Year 2021	1 Jan - 31 Dec	365	688.17	[21.5, 256]
W1	1 Jan - 4 Sep	222	368.92	[1.73, 127]
W2	5 Sep - 9 Nov	66	2233.53	[6.84, 119.97]
W3	10 Nov - 31 Dec	77	282.89	[1.15, 25.49]

Subsequently, the district-specific disease incidence rates calculated for each temporal period using Equation 4.1.

$$\text{Incidence rate} = \left(\frac{\text{the number of confirmed cases}}{\text{the population of a district}} \right) \times 1000 \quad \text{Equation 4.1}$$

The change point analysis of COVID-19 data in Sarawak, Malaysia, reveals important spreading dynamics of the disease throughout 2021. During each identified period (W1, W2, W3), the average daily cases offer insights into the intensity of COVID-19 transmission daily.

Throughout the year, Sarawak experienced an average of approximately 688.17 daily cases, with district-wise incidence rates ranging from 21.5 to 256 cases per 1000 population. In the First wave (January 1st to September 4th), the average daily cases were around 368.92, indicating moderate transmission levels. The second wave (September 5th to November 9th) saw a significant surge in cases, with an average of 2233.53 daily cases. This sharp increase in COVID-19 cases is noteworthy and could be attributed to changes in virus variants, large gatherings, relaxed public health measures, and increased community transmission. The third wave (November 10th to December 31st) showed a decline in daily cases to approximately 282.89, suggesting a decrease in disease activity and transmission compared to wave 2.

The histogram in Figure 4.3 gives a graphical representation of the distribution of COVID-19 cases in Sarawak, Malaysia. It illustrates the frequency of daily case counts, showing how many days recorded a certain number of cases. This visualization allows us to identify the normality patterns and assess the overall trend of the pandemic's daily case count over 2021 and within its three distinct waves. By examining the histogram, this study can

gain insights into the variations in disease incidence and understand the high and low transmission intensity periods.

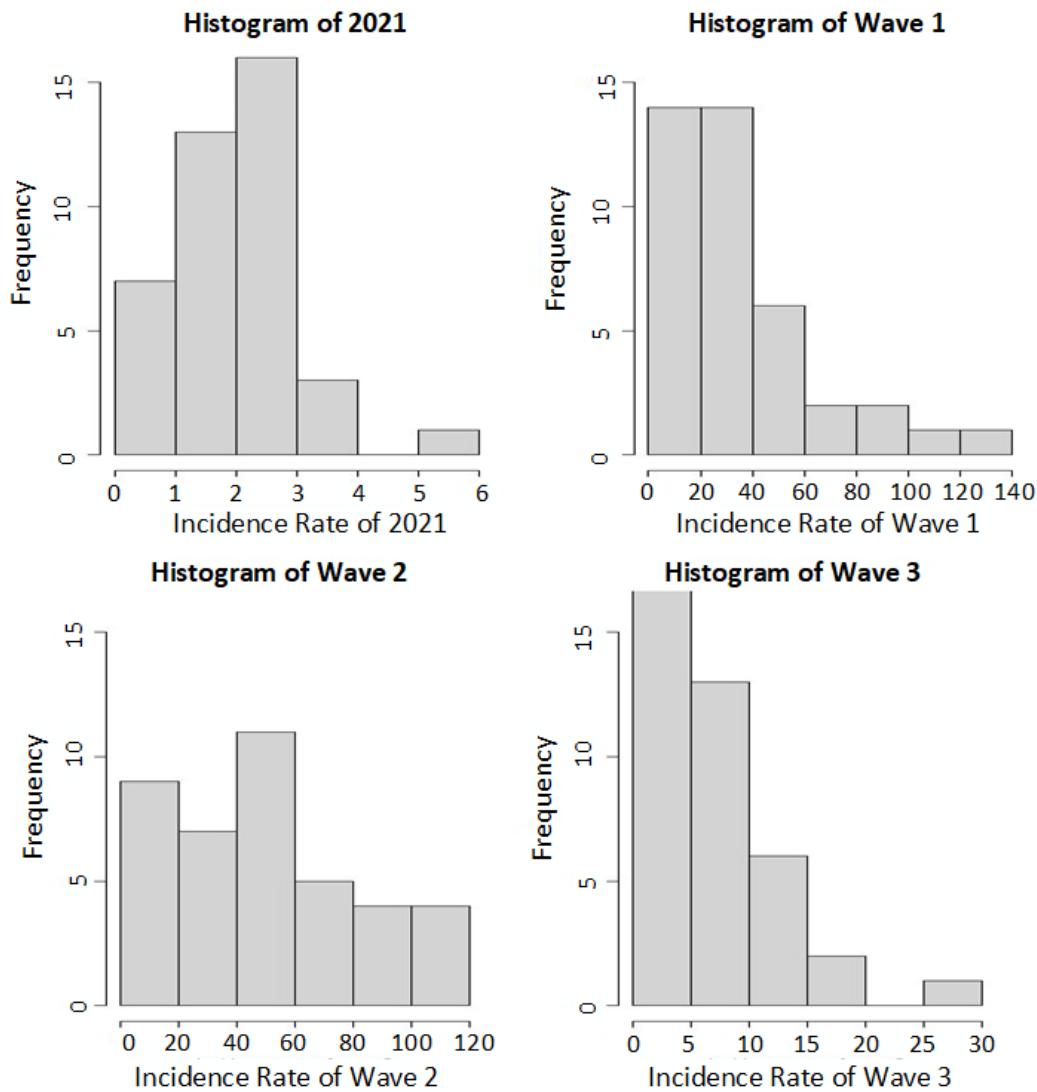


Figure 4.3: The histograms for incidence rates in different waves and 2021

Each histogram provides valuable insights into the spread of COVID-19 and the intensity of cases during these specific time frames. The histograms show the incidence rates on the x-axis, representing the average number of new daily cases. In contrast, the y-axis represents the frequency or count of districts with similar incidence rates.

The histogram helps reveal the data distribution with a wide range of skewness. As the histograms in Figure 4.3 are right-skewed, logarithm transformation is applied. This

transformation It compresses values, reduces outliers' impact, and uncovers patterns. This aids in meeting normality assumptions, facilitating statistical analysis, and drawing meaningful conclusions in data analysis.

Figure 4.4 give the histograms of the log-transformed data for different variables. In the first plot, this study examined the log-transformed population size. The histogram shows that the distribution of the log-transformed population size can be considered following a negative exponential distribution across the districts in Sarawak, where several urban district population sizes are relatively high compared to most of the rural districts.

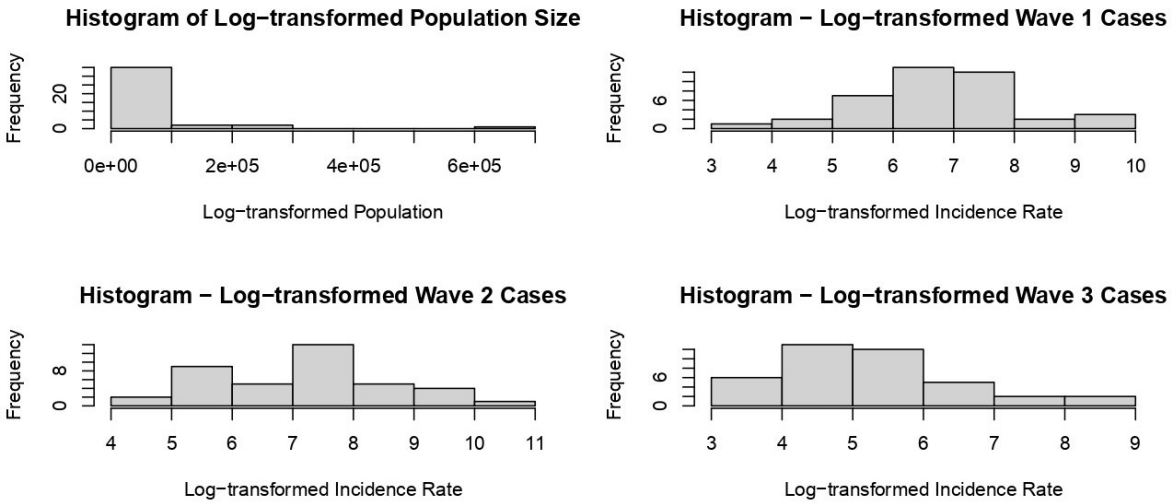


Figure 4.4: Histograms of log-transformed population size and COVID-19 cases for Wave 1, Wave 2, and Wave 3 in Sarawak districts

The subsequent histograms in Figure 4.4 represent the log-transformed data for each of the three waves. These plots ensure that the data is approximately normally distributed and suitable for further statistical analysis and modeling. Log transformation is often employed to handle skewed data and to meet the assumptions of specific statistical techniques that usually require normally distributed data.

4.3 Constructing Spatial Weights Matrix in R Programming

Several operations on the adjacency-based Spatial Weights matrix are briefly described as follows. The adjacency matrix is imported from a CSV file and transformed into a matrix class. The row names of the matrix are assigned as column names to ensure proper labeling. Subsequently, the districts are organized alphabetically. The matrix is then row-standardized by dividing each value with the sum of weights in its corresponding row. This standardization ensures the weights reflect relative proportions and enables meaningful district comparisons. The row sums of the standardized matrix are checked to verify that they are equal 1 to validate the scaling.

The next step involves converting the matrix into a weights list object using the ``mat2listw ()`` function from the ``spdep`` library. This transformation facilitates further spatial analysis and modeling of the interconnections among districts. Key summaries are generated to gain insights into the weights list object's characteristics. These include the number of districts (40), the count of nonzero links (170), the percentage of nonzero weights (10.625%), and the average number of links (4.25) per district. Moreover, the distribution of link numbers is examined; Lawas is identified as the least connected district with one link and Sibuh as the most connected district with eight links.

Finally, the weights distribution is summarized, providing information on the number of districts, the total number of links, and the sums of weights, absolute weights, and squared weights. These operations are essential for comprehending the spatial relationships and dependencies among the 40 districts. This understanding is crucial for conducting meaningful spatial analysis and modeling tasks, enabling valuable insights for various fields, such as geography, urban planning, and epidemiology.

After performing the described operations on the adjacency-based Spatial Weights matrix, the bar plot complements the spatial analysis, providing a concise and insightful summary of the districts' spatial relationships.

The bar plot (Figure 4.5) displays the distribution of link numbers among districts. The x-axis represents the number of links, ranging from 1 to 8, while the y-axis shows the frequency or count of the districts with their link number. The plot provides valuable insights into the connectivity patterns among the districts. A few districts have only 1 or 8 links, as denoted by the lower frequencies at the edges of the distribution.

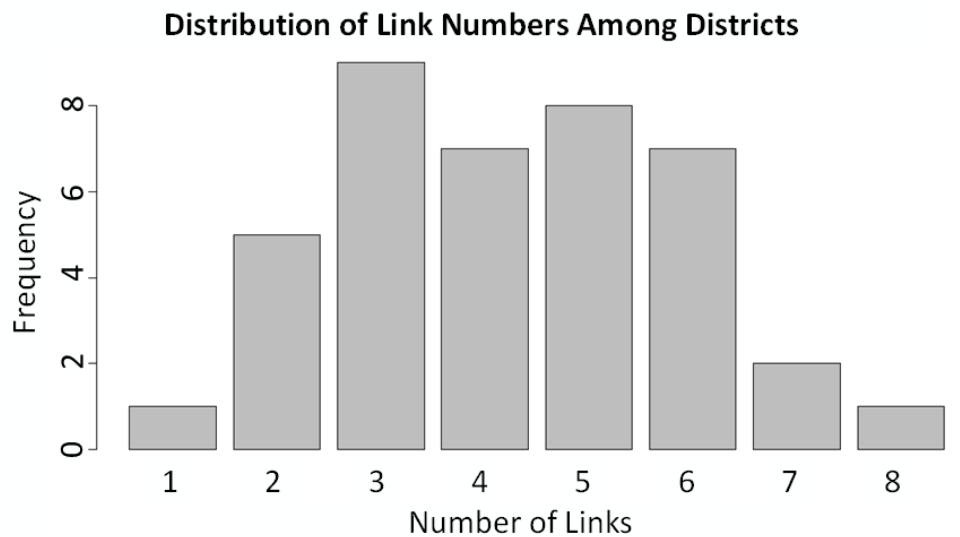


Figure 4.5: Bar plot that gives the distribution of link numbers among district

In Figure 4.6, the `igraph` package is utilized to create a graph object representing the adjacency matrix `W_40_Districts`, which illustrates the spatial connections between 40 districts. The graph is plotted with customized settings, showing the district names as vertex labels for easy identification. The "undirected" mode ensures bidirectional connections between districts. The linked edge indicates the weights between districts in the adjacency matrix, showcasing the spatial relationships among them. For instance, Limbang is connected by only one edge to Marudi, whereas Song district has six edges connecting it to

six different districts. This implies that Song has better spatial connectivity compared to Limbang. This visualization offers insights into spatial clustering, central districts, and isolated areas, aiding in understanding spatial interactions among districts. The graph serves as a valuable tool for spatial analysis, enabling the identification of patterns and further exploration of the spatial structure and relationships among the districts.

Figure 4.6: The schematic diagram for illustrating spatial connection between districts

The global spatial autocorrelation analysis using first-order queen contiguity spatial Weights was performed on the COVID-19 incidence rates in Sarawak for three different periods. All Moran's I statistics calculated in Table 4.3 indicate that its incidence rate tends

to clustering (i.e., positive spatial autocorrelation), and they are statistically significant with p -values < 0.05 . The findings suggest that the incidence rates among districts that are geographically adjacent or near are closer if compared to the districts that are relatively far.

Table 4.3: Global spatial autocorrelation analysis using first-order queen contiguity spatial weights

	Incidence rate			
	2021	W1	W2	W3
Moran's I statistic	5.411	5.975	7.726	5.335
<i>P</i> -value	3.118e-08	1.144e-09	5.530e-15	4.774e-08

The Moran's I statistics for each period were 5.41, 5.98, 7.73, and 5.34, respectively, all statistically significant with very small p -values (close to 0). It suggests that the observed spatial clustering of COVID-19 cases in Sarawak during 2021 and the individual waves is highly unlikely to have occurred by chance, indicating the presence of meaningful spatial patterns influencing the spread of the virus across the district.

The global spatial autocorrelation analysis revealed significant spatial clustering, indicating that districts with high incidence rates were spatially connected to neighboring districts with similarly high rates. In contrast, low-rate districts were adjacent to other low-rate districts. This clustering highlighted the role of spatial interactions in the spread of COVID-19. This study also generated Moran's scatter plot to delve deeper into the relationship between connectivity and spatial autocorrelation.

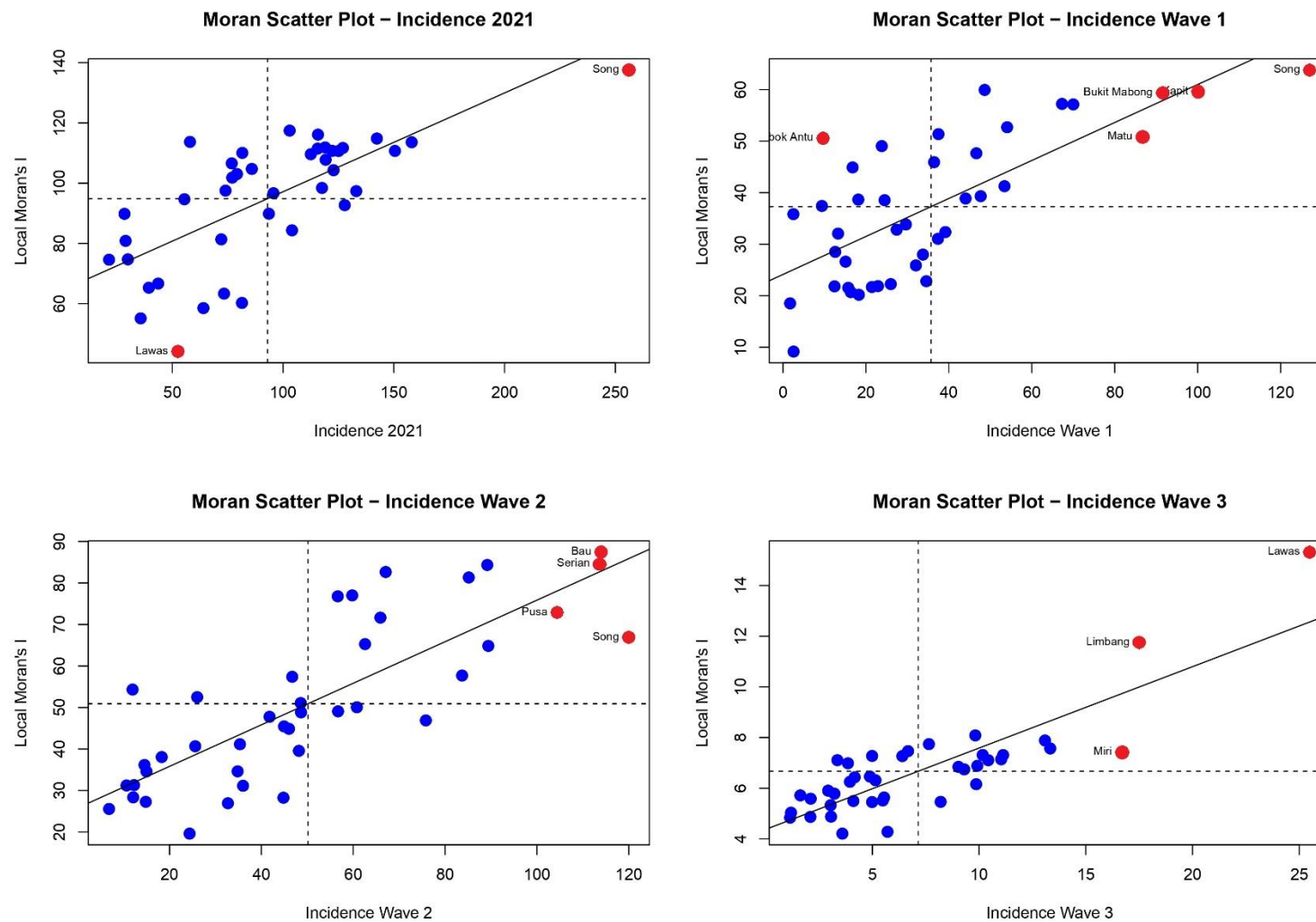


Figure 4.7: Moran's I scatter plot

The scatter plots show the relationship between district-wise COVID-19 incidence rates in 2021 (respectively, wave 1, 2, and 3) and the local Moran's I values, illustrating their spatially lagged values (Figure 4.7). The *x*-axis represents the incidence rates, while the *y*-axis indicates the local Moran's I values. Red dots indicate significant spatial autocorrelation, while blue dots represent non-significant cases. The plots for each wave reveal changes in spatial patterns over time. The `moran.plot` function constructs these Moran Scatter Plots, illustrating the spatial clustering of incidence rates across districts.

4.5 Local Indicators of Spatial Association (LISA)

The LISA cluster map, generated using first-order queen contiguity spatial weights, reveals local spatial autocorrelation patterns. The map highlights districts with significant positive or negative spatial autocorrelation through color coding. Districts in red and blue indicate spatial clusters where neighboring districts exhibit similar incidence rates, suggesting high or low incidence rate concentrations. Conversely, districts in paler colors, such as pink and light blue, denote negative spatial autocorrelation, indicating spatial outliers surrounded by neighbors with contrasting incidence rates (Figure 4.8). Overall, the map visually depicts local spatial relationships and helps identify incidence rate patterns among districts.

In Figure 4.8, the red-colored districts highlight significant spatial clusters with high incidence rates. Throughout the year, these high-high districts included Sri Aman, Kanowit, and Song (see Fig 4.8(a)), and during the first waves, clusters were seen in central Sarawak districts such as Bukit Mabong, Kapit, Tatau, Mukah, Selangau, Sibu, Meradong, Julau, Kanowit, and Song (see Fig 4.8(b)). In the second wave, these clusters shifted to South-West Sarawak, covering Lundu, Bau, Kuching, Samarahan, Serian, Tebedu, Asajaya, Simunjan,

Sri Aman, Pusa, Betong, and Song (see Fig 4.8(c)). During the third wave, only Limbang and Lawas in North-East Sarawak remained in the high-high cluster (see Fig 4.8(d)).

The analysis also identified some low-high clusters in the LISA map for the entire year of 2021 (see Fig 4.8(a)), and several low-low clusters, mainly in northern districts such as Miri, Beluru, Telang Usan, and Sebauh, and in other parts of Sarawak, including Daro and Matu during the second wave (see Fig 4.8(c)), and Bukit Mabong in the south during the third wave (see Fig 4.8(d)). These clusters are considered potential disease transmitters.

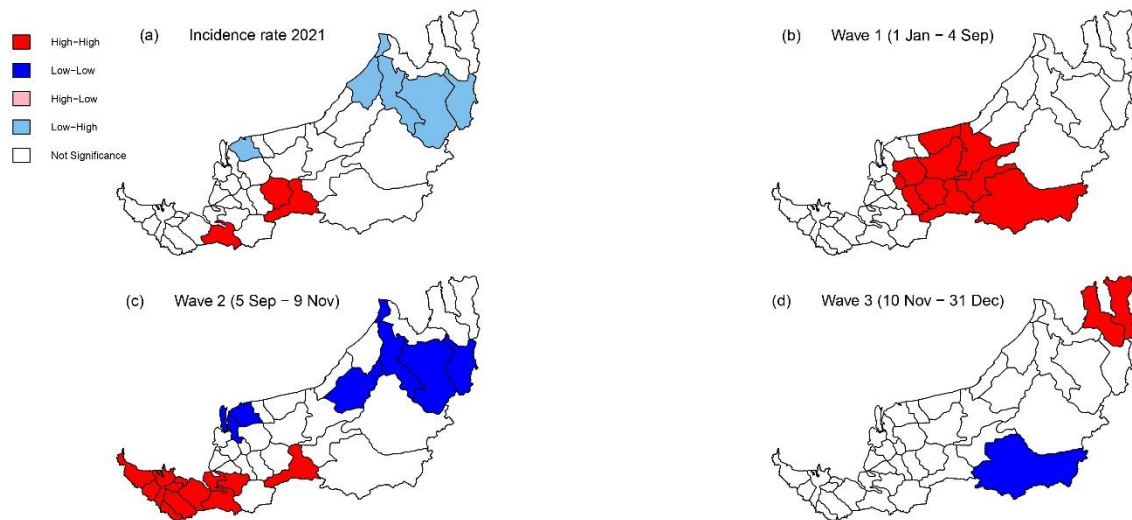


Figure 4.8: Local indicator of spatial association (LISA) cluster map for incidence rates in different periods

4.6 Global Non-spatial / Spatial Regression Models (OLS, SLM, and SEM)

The result of each regression model is given in this section. Tables 4.4 to 4.6 summarize the statistical significance of the coefficients in four different regression models (Incidence 2021, Incidence W1, Incidence W2, and Incidence W3) with various socio-demographic variables. The significance levels are denoted by asterisks (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$), indicating the probability of obtaining the observed results by chance.

Table 4.4: The coefficient estimates and statistical test values for OLS

	Incidence 2021	Incidence W1	Incidence W2	Incidence W3
Pop GrowthR	-0.013	0.046	-0.079	-0.210*
Pop NonCitizen	-0.305**	-0.233	-0.292	-0.134
Pop Male	0.13	0.197	0.117	-0.028
Pop 15to64	-0.259	-0.269	-0.109	-0.400**
Pop 65	0.005	0.155	0.127	-0.057
Pop Density	0.12	0.187	0.195	-0.102
Internet	-0.257	-0.502	-0.306	0.252
Poverty	-0.101	-0.136	-0.032	0.217
Gini	-0.009	-0.186	-0.025	0.165
Median inc	0.103	-0.144	0.259	0.366
Household size	-0.058	-0.176	0.057	-0.257*
Garbage	0.129	0.802**	-0.1	0.042
Piped water	0.247	0.211	0.323	0.214
Public h	-0.267	-0.700**	-0.081	-0.229
Constant (Intercept)	4.403***	3.231***	3.680***	1.729***
Parameters				
SSE	6.66	16.34	14.80	7.88
Akaike Inf. Crit.	73.80	109.72	105.75	80.57
R^2	0.428	0.538	0.329	0.614
Adjusted R^2	0.108	0.279	0.047	0.398
Residual Std. Error (df = 25)	0.516	0.809	0.77	0.562
F Statistic (df = 14; 25)	1.337	2.078	0.876	2.839
RMSE	0.408	0.640	0.609	0.444
Value of p	* $p < 0.1$;	** $p < 0.05$;	*** $p < 0.01$	

For OLS model (see Table 4.4), the population variables (Pop GrowthR, Pop Noncitizen, Pop 15to64) and household size show some statistically significant associations with the incidence rate at different temporal periods. Additionally, access to piped water (Piped water) is positively associated, while internet access (Internet), poverty, and the Gini coefficient are negatively associated with the incidence rate, except in Wave 3. For Incidence W1, Household size and garbage disposal (Garbage) access are statistically significant predictors of incidence rate during Wave 1. For Incidence W2, no variables are statistically

significant, and the direction of association with the incidence rate changes for some factors. For Incidence W3, the model has more variables that are statistically significant. Median income and access to public healthcare facilities are positively associated with the incidence rate, while access to garbage disposal shows a negative association. Overall, the significance levels and the patterns of association between the incidence rate and socio-demographic variables vary across different waves, highlighting the potential variations in factors affecting COVID-19 incidence during different periods.

In the analysis of the SLM model (see Table 4.5), the incidence rate shows statistically significant associations with several population variables (Pop GrowthR, Pop NonCitizen, Pop 15to64, Pop 65) and household size. Additionally, access to piped water is positively associated, while internet access (Internet), poverty, and the Gini coefficient (Gini) are negatively associated with the incidence rate except for Wave 3. Similar patterns are observed in Incidence W1, with some additional variables showing significance. Specifically, household size (Household size) and access to garbage disposal (Garbage) are statistically significant predictors of the incidence rate during Wave 1. However, in Incidence W2, fewer variables are statistically significant, and the direction of association with the incidence rate changes for some factors. Notably, access to public healthcare facilities (Public h) and internet access (Internet) are negatively associated with the incidence rate during Wave 2.

Table 4.5: The coefficient estimates and statistical test values for SLM based on first-order queen contiguity

	Incidence 2021	Incidence W1	Incidence W2	Incidence W3
Pop GrowthR	-0.015	0.049	-0.097	-0.211**
Pop NonCitizen	-0.302***	-0.248	-0.263*	-0.131
Pop Male	0.133	0.155	0.155	-0.017
Pop 15to64	-0.263**	-0.261	-0.163	- 0.406***
Pop 65	-0.029	0.088	-0.046	-0.059
Pop Density	0.101	0.162	0.075	-0.097
Internet	-0.262*	-0.374	-0.429**	0.23
Poverty	-0.098	-0.123	-0.043	0.212*
Gini	-0.0002	-0.142	0.019	0.159*
Median inc	0.102	-0.165	0.282	0.376*
Household size	-0.064	-0.173	0.001	-0.256**
Garbage	0.133	0.737***	-0.062	0.051
Piped water	0.269*	0.173	0.484**	0.226
Public h	-0.260**	- 0.669***	-0.038	-0.234
Constant (Intercept)	3.867***	2.562***	2.035***	1.620***
Parameters				
SSE	6.50	15.62	12.32	7.78
Akaike Inf. Crit.	75.477	110.418	103.745	82.479
R^2	0.945	0.869	0.897	0.935
Adjusted R^2	0.870	0.687	0.753	0.844
Residual Std. Error (df = 25)	0.768	1.191	1.058	0.840
F Statistic (df = 14; 25)	0.768	5.654	7.395	12.205
RMSE	0.543	0.842	0.748	0.594
P value	* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$			

In contrast, Incidence W3 includes more statistically significant predictors. Median income (Median inc) and access to public healthcare facilities (Public h) are positively associated with the incidence rate during Wave 3. In contrast, access to garbage disposal (Garbage) shows a negative association. These results highlight the dynamic nature of the

factors influencing COVID-19 incidence during different periods, indicating potential variations in the drivers of transmission over time.

Based on the summary of Table 4.5, Incidence w1 stands out as the better model among Incidence 2021, Incidence W2, and Incidence W3. Incidence W3 demonstrates the highest number of significant predictor variables, as the *p*-values indicate. For several vital predictors, such as Pop GrowthR, Pop NonCitizen, Pop 15to64, Internet, Household size, and Piped water, Incidence w1 shows statistically significant relationships with the incidence rate. It suggests that Incidence w1 provides a more robust and meaningful representation of the factors influencing the incidence rate compared to the other models. Furthermore, Incidence w1 exhibits the highest AIC value, indicating a better fit to the data and better explanatory power. Therefore, Incidence w1 is the preferred model as it captures more significant predictors and yields a more parsimonious and accurate representation of the incidence rate.

The analysis of the SEM model (see Table 4.6) reveals varying patterns of significance for different socio-demographic variables. In Incidence 2021, the incidence rate exhibits statistically significant associations with population variables such as Pop NonCitizen, Pop 15to64, and Pop Density, indicating that population dynamics play a role in COVID-19 transmission. Household size also shows significance. Similar patterns are observed as this study move to Incidence w1, with additional variables, including Garbage, showing significant associations. However, in Incidence W2, fewer variables are statistically significant, and the direction of association with the incidence rate changes for some factors. Notably, access to public healthcare facilities (Public h) and internet access (Internet) are negatively associated with the incidence rate during Wave 2, indicating potential shifts in transmission determinants.

Table 4.6: The coefficient estimates and statistical test values for SEM based on first-order queen contiguity

	Incidence 2021	Incidence W1	Incidence W2	Incidence W3
Pop GrowthR	0.009	0.044	0.008	-0.246***
Pop NonCitizen	-0.320***	-0.238	-0.303**	-0.202**
Pop Male	0.125	0.217	0.122	-0.075
Pop 15to64	-0.279**	-0.267	-0.128	-0.426***
Pop 65	0.143	0.172	0.398***	-0.082
Pop Density	0.212***	0.219	0.303***	-0.131
Internet	-0.235*	-0.580**	-0.175	0.447***
Poverty	-0.043	-0.114	0.009	0.265***
Gini	-0.083	-0.221*	-0.156*	0.251***
Median inc	0.232	-0.122	0.469**	0.261*
Household size	-0.048	-0.189	0.177	-0.292***
Garbage	0.127	0.853***	-0.097	-0.004
Piped water	0.151	0.241	0.041	0.165
Public h	-0.424***	-0.756***	-0.226	-0.397***
Constant	4.418***	3.233***	3.705***	1.764***
Parameter				
SSE	3.98	14.73	7.90	4.38
Akaike Inf. Crit.	71.544	111.546	101.927	74.234
R^2	0.966	0.877	0.934	0.963
Adjusted R^2	0.920	0.705	0.842	0.912
Residual Std. Error (df = 25)	0.601	1.157	0.847	0.631
F Statistic (df = 14; 25)	24.666	6.047	12.006	22.336
RMSE	0.425	0.818,	0.599	0.446
Value of p	* $p<0.1$; ** $p<0.05$; *** $p<0.01$			

In contrast, Incidence W3 includes more variables that are statistically significant predictors, such as Median inc and access to public healthcare facilities (Public h), showing positive associations with the incidence rate. At the same time, access to garbage disposal (Garbage) exhibits a negative association. Overall, the analysis demonstrates dynamic changes in the significance of socio-demographic factors across different waves, highlighting the evolving nature of COVID-19 transmission drivers over time.

Based on the summary of the significance values, Incidence w3 emerges as the superior model among Incidence 2021, Incidence W2, and Incidence W3. Incidence W3 demonstrates the highest number of statistically significant predictor variables, including Pop GrowthR, Pop 15to64, Internet, Poverty, Gini, and Median inc, indicating that it effectively captures the influence of these factors on the incidence rate. It suggests that Incidence w3 provides a more comprehensive and meaningful representation of the factors influencing the incidence rate compared to the other models. Additionally, Incidence W3 exhibits the lowest AIC value, implying a better fit to the data and higher explanatory power. Therefore, Incidence W3 is the preferred model as it captures more significant predictors and yields a more parsimonious and accurate representation of the incidence rate.

Upon examining various models (including OLS, SLM, and SEM), it is evident that the model varies depending on the specific dependent variable. OLS falls within the non-spatial category, while SLM and SEM are categorized as spatial models. In this context, Wu and Zhang (2013) suggested that one can take into account the AIC and the SSE. Table 4.7 reveals that SEM consistently displays the lowest AIC and SSE values compared to the other two models for the incidence in 2021. The same holds for Incidence w2 and Incidence w3, where SEM exhibits strong performance. Only for Incidence w1, mixed results are observed, with OLS and SEM having the lowest AIC and SSE values, respectively. Nonetheless, the disparities between OLS and SEM are minimal, and SEM performs well for the remaining waves and within the same wave in terms of SSE. In summary, SEM outperforms the other two models.

Table 4.7: AIC and SSE results of all three models for all waves

		Incidence 2021	Incidence W1	Incidence W2	Incidence W3
AIC	OLS	73.80	109.72	105.75	80.57
	SLM	75.19	110.50	103.14	82.20
	SEM	71.54	111.54	101.92	74.23
SSE	OLS	6.66	16.34	14.80	7.88
	SLM	75.47	110.41	103.74	82.47
	SEM	3.98	14.73	7.90	4.38

The results of the OLS, SLM, and SEM models were tested on data from the entire year of 2021 and on segmented data representing different waves (which can consider as validation data). Overall, the SEM model demonstrated consistent results. The models' performances were evaluated using various parameters, including SSE, Akaike Inf. Crit., R^2 , Adjusted R^2 , Residual Std. Error (df = 25), F Statistic (df = 14; 25), RMSE, to determine the best model. It was observed that the effectiveness of each model varied depending on the specific dependent variable, with OLS being a non-spatial model and SLM and SEM being spatial models. To resolve the differences, this study focused on AIC and SSE for final conclusions, finding that SSE was a reliable performance indicator. Based on this analysis, the study concluded that the tested models and parameters can be effectively used for forecasting future COVID-19 confirmed cases.

4.7 Summary

The primary objective of this study was to perform a spatial auto-correlation analysis of COVID-19 cases in Sarawak's districts. The steps involved include collecting socio-demographic data and COVID-19 case data, then segmenting the year 2021 into three waves using change-point analysis. The timelines identified were: the entire year (January 1 - December 31), wave 1 (January 1 - September 4), wave 2 (September 5 - November 9), and wave 3 (November 10 - December 31). Sarawak experienced an average of approximately

688.17 daily cases in 2021, with district-wise incidence rates ranging from 21.5 to 256 cases per 1000 population. Wave 1 had an average of 368.92 daily cases, wave 2 saw a surge to 2233.53 daily cases, and wave 3 had 282.89 daily cases, indicating a decline compared to wave 2. The study employed Moran's I statistics and its p -value to examine incidence rates using first-order queen contiguity spatial weights. The findings indicated that geographically adjacent districts had more similar incidence rates compared to those further apart. Positive spatial autocorrelation (clustering) was observed in all periods, with statistically significant Moran's I statistics of 5.41, 5.98, 7.73, and 5.34, suggesting meaningful spatial patterns influencing the spread of COVID-19 across districts.

LISA maps highlighted significant spatial clusters throughout the year, including Sri Aman, Kanowit, and Song. In wave 1, clusters were also seen in central Sarawak districts such as Bukit Mabong, Kapit, Tatau, Mukah, Selangau, Sibul, Meradong, Julau, and Song. During wave 2, clusters shifted to South-West Sarawak, covering Lundu, Bau, Kuching, Samarahan, Serian, Tebedu, Asajaya, Simunjan, Sri Aman, Pusa, Betong, and Song. In wave 3, only Limbang and Lawas in North-East Sarawak remained in the high-high cluster. The analysis also identified low-high and low-low clusters in various districts, which are considered potential disease transmitters.

Finally, global non-spatial and spatial regression models (OLS, SLM, and SEM) were applied to the data from 2021 and the three waves. The model performance was evaluated using the AIC and the SSE. SEM consistently displayed the lowest AIC and SSE values, outperforming OLS and SLM for the incidence in 2021 and in waves 2 and 3. Mixed results were observed for wave 1, with both OLS and SEM showing strong performance. Overall, SEM was found to be the most effective model for analyzing the COVID-19 incidence rates in Sarawak.

CHAPTER 5

SPATIAL AUTO-CORRELATION ANALYSIS ON STATE-WISE AND DISTRICT-WISE INCIDENCE USING SUPER ADJACENCY MATRIX

This chapter focuses on identifying spatial clusters of high or low COVID-19 incidence rates in Malaysia, examining the state of Malaysia, the district of Sarawak, and the impact of travel on spreading COVID-19 from state to district. Moran's I values are used to identify these clusters, with a thorough examination of the weight matrix for both state and two-level analyses. The supra adjacency method introduces the weight matrix for state-to-district connections. Additionally, the degree, betweenness, and eigenvector centrality for flight routes from state to district, state to state, and district to district are assessed.

LISA diagrams are created for three scenarios: the entire state of Malaysia, a two-level analysis, and the state-district level. The primary objective is to investigate spatial clusters using incidence rates in three contexts: the state-wise impact on Malaysia, the combined impact of the state of Malaysia and the district of Sarawak, and the influence of the state of Malaysia on the spread of COVID-19 within Sarawak, considering travel patterns in 2021. The flow diagram of this chapter is presented in Figure 5.1.

5.1 Local Moran's I for State, Two-Level and State to District

Local Moran's I is a spatial statistic used in spatial analysis to identify clusters or patterns of spatial autocorrelation in the incidence rate of COVID-19 (Li et al., 2020). It measures whether a particular feature at a specific location resembles the surrounding features. In other words, it helps identify local spatial clusters of high values near other high values (high-high clusters) or low values near other low values (low-low clusters). This study considered 14 features and details already mentioned in Chapter 4.

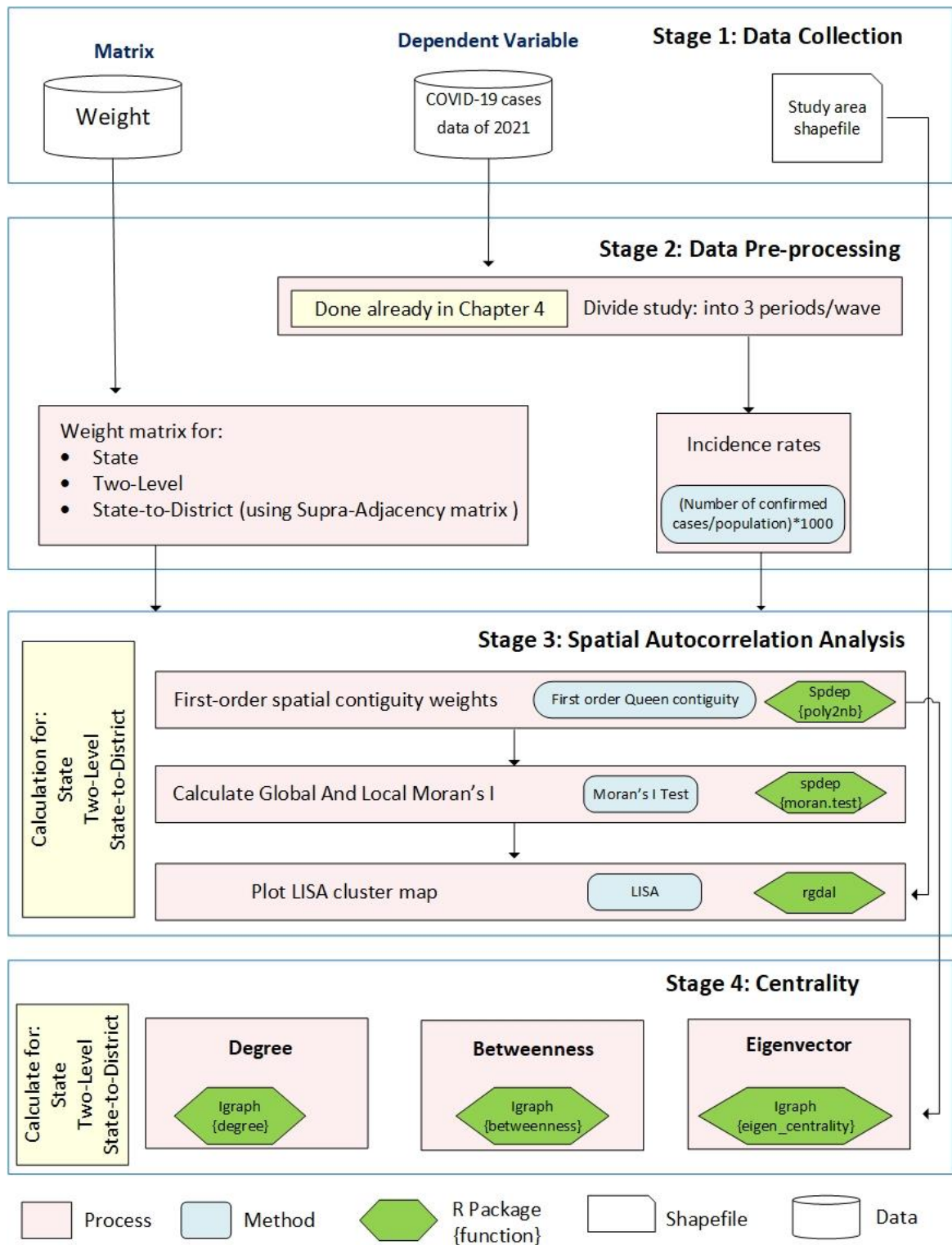


Figure 5.1: Flow diagram of chapter 5

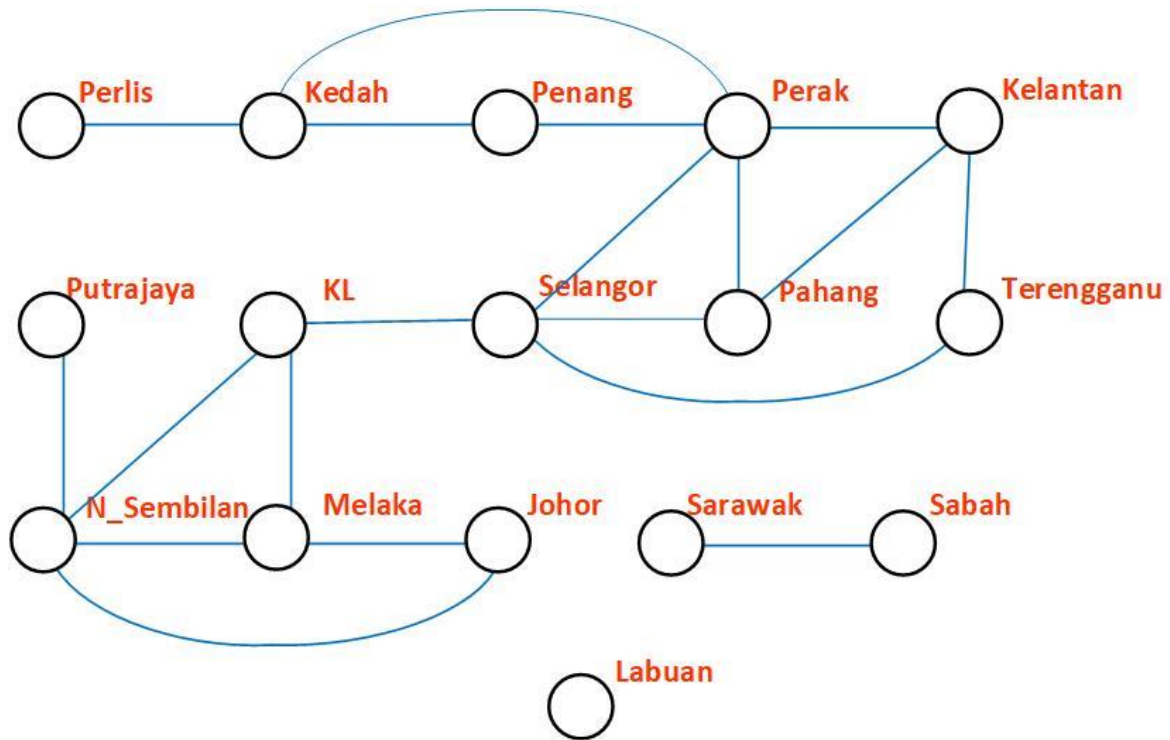


Figure 5.2: Plot of weight matrix of State

5.1.1 LISA for State of Malaysia

As mentioned in Chapter 4 (page no 38-40), incidence rate weight is required to calculate Moran's I value. Thus, this study calculates the incidence rate and weight before calculating Moran's I.

Malaysia comprises 13 states and three federal territories: Kuala Lumpur, Putrajaya, and Labuan. For this study, these three federal territories are treated as states, effectively considering 16 states in Malaysia. The weight matrix for these 16 states is depicted in Figure 5.2. Within Figure, the blue-colored dots symbolize a value of 1, while empty cells signify a value of 0.

The incidence rate represents the count of new disease cases within a specific population during a defined time frame. It is computed by dividing the number of new cases by the total population and multiplying the result by the number of days in the specified period. The formula for calculating the incidence rate is as follows:

$$\text{Incidence rate} = \left(\frac{\text{Number of new cases}}{\text{Total population}} \right) \times \text{Number of days} \quad \text{Equation 5.1}$$

The analysis is focused on the entire dataset from the year 2021, which is further divided into three waves based on COVID-19 case data. This study assumes that the duration and number of COVID-19 waves in Sarawak align with the state of Malaysia to allow for a transparent comparison and to assess the impact of the state of Malaysia on transmitting the virus to Sarawak districts. Consequently, this study does not recalculate the number and duration of waves for the state of Malaysia. More information regarding these three waves can be found in Chapter 4, section 4.2 (page no 49). The data for new cases and total population were obtained from the official website of the Ministry of Health of Malaysia.

Table 5.1 shows the incidence rate of COVID-19 in each state of Malaysia in 2021 and the incidence rate in each wave of the pandemic. The incidence rate is the number of new cases per 1,000 people. Furthermore, the highest incidence rate was in Selangor, followed by Kuala Lumpur, Negeri, Sembilan, and the lowest incidence rate of COVID-19 was in Perlis, Perak, and Sabah in 2021. Among all the waves, Labuan and Perlis in Wave 1, Sarawak and Labuan in Wave 2, and Putrajaya and Perak in Wave 3 has the highest and lowest incidence rate, respectively. It can be observed that Labuan has the highest incidence rate in Wave 1 and the lowest in Wave 2. However, Sarawak has the highest incidence rate in Wave 2, which justifies the state that Sarawak witnessed the COVID-19 cases in Wave 2.

The information in Table 5.1 can be used to understand the incidence rate of COVID-19 in Malaysia and to develop more effective interventions to control the spread of the disease. With the help of incidence rate and weight, Local Moran's I values are calculated, and based on that, a LISA map is drawn.

Table 5.1: Incidence rate of Sates of Malaysia for the year and all three waves

No.	State	Incidence 2021	Incidence W1	Incidence W2	Incidence W3
1	Perlis	27.878	3.178	15.757	8.943
2	Kedah	75.383	25.862	37.800	11.720
3	Penang	88.837	28.481	48.081	12.274
4	Perak	50.157	16.985	25.975	7.195
5	Kelantan	87.244	26.242	40.812	20.189
6	Terengganu	64.582	16.664	33.583	14.334
7	Pahang	55.785	17.156	24.918	13.710
8	Selangor	115.060	67.992	33.325	13.742
9	Kuala Lumpur	112.998	74.920	25.942	12.135
10	Putrajaya	78.441	37.364	17.037	24.039
11	Negeri Sembilan	93.662	64.423	17.163	12.075
12	Melaka	80.043	37.638	27.083	15.322
13	Johor	63.045	25.642	27.957	9.444
14	Sarawak	88.923	28.900	52.038	7.984
15	Sabah	53.096	15.505	29.341	8.249
16	Labuan	91.138	80.109	2.747	8.281

This study focused on 2021 incidence rate data to evaluate the LISA diagram for the 2021 year and all three waves (Figure 5.3).

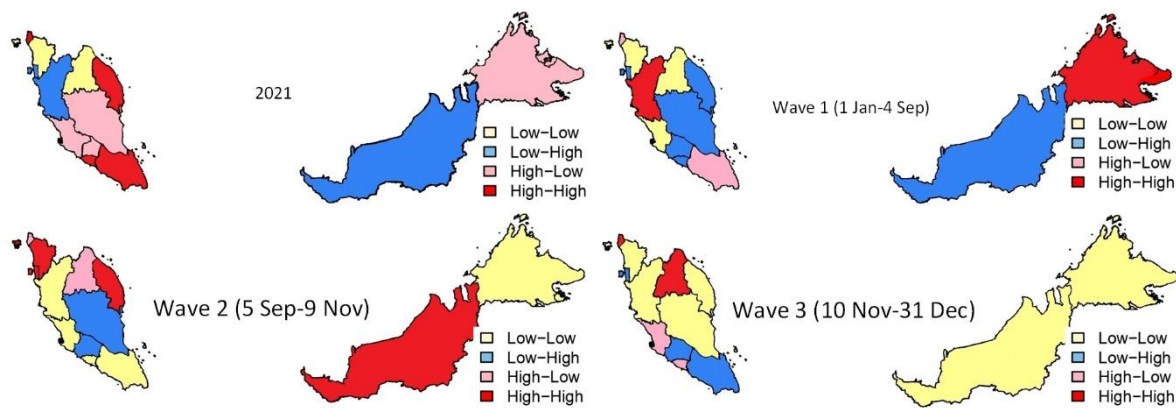


Figure 5.3: LISA diagram of state of Malaysia

The LISA map indicates that in 2021, the COVID-19 incidence rate was notably high in various regions of Malaysia throughout the year and during different waves. For the entire year, high-high clusters were identified in Perlis, Terengganu, Melaka, and Johor, all located

in Peninsular Malaysia. When examining wave-specific data, Perak and Sabah were in the high-high cluster during wave 1, Kedah, Terengganu, and Sarawak during wave 2, Perlis and Kelantan during wave 3.

Additionally, in 2021, Pahang, Selangor, and Negeri Sembilan from West Malaysia, along with Sabah from East Malaysia, were in the high-low cluster. Perak and Sarawak were in the low-high cluster for the year 2021. In 2021, Perak, Penang, and Sarawak were in the low-high cluster, while Terengganu, Pahang, Negeri Sembilan, Melaka and Sarawak were in low-high cluster in wave 1. Pahang and Negeri Sembilan remained in the low-high cluster during wave 2, and Negeri Sembilan and Johor were in wave 3.

As previously noted, it is essential to highlight that the highest incidence rate during wave 2 occurred in the Sarawak state, which also witnessed the highest number of COVID-19 cases in a relatively short time. Consequently, to account for these dynamics, the analysis considered the state of Malaysia and the districts of Sarawak in a two-level approach.

5.1.2 LISA for Two-level

Two level represents the combination of the state of Malaysia and the district of Sarawak state. A Shapefile, weight matrix, and incidence rate are required to draw a LISA map. Shapefile for the state and Sarawak district is available. However, this study needs a shape file for all states of Malaysia and all districts of Sarawak. Thus, this study merged the shape file of all states of Malaysia and the district of Sarawak state. For the weight matrix of two levels, it is required to combine the matrix of state and district. However, the issue is that the size of the state matrix is 16×16 , and the district is 40×40 ; hence, this study need to add a null matrix to combine it using Equation 5.2. Therefore, the weight matrix for the two-level is 56×56 .

$$\text{Two-level weight matrix} = \begin{bmatrix} \text{State}_{16 \times 16} & \text{Null Matrix}_{16 \times 40} \\ \text{Null Matrix}_{40 \times 16} & \text{Districts}_{40 \times 40} \end{bmatrix} \quad \text{Equation 5.2}$$

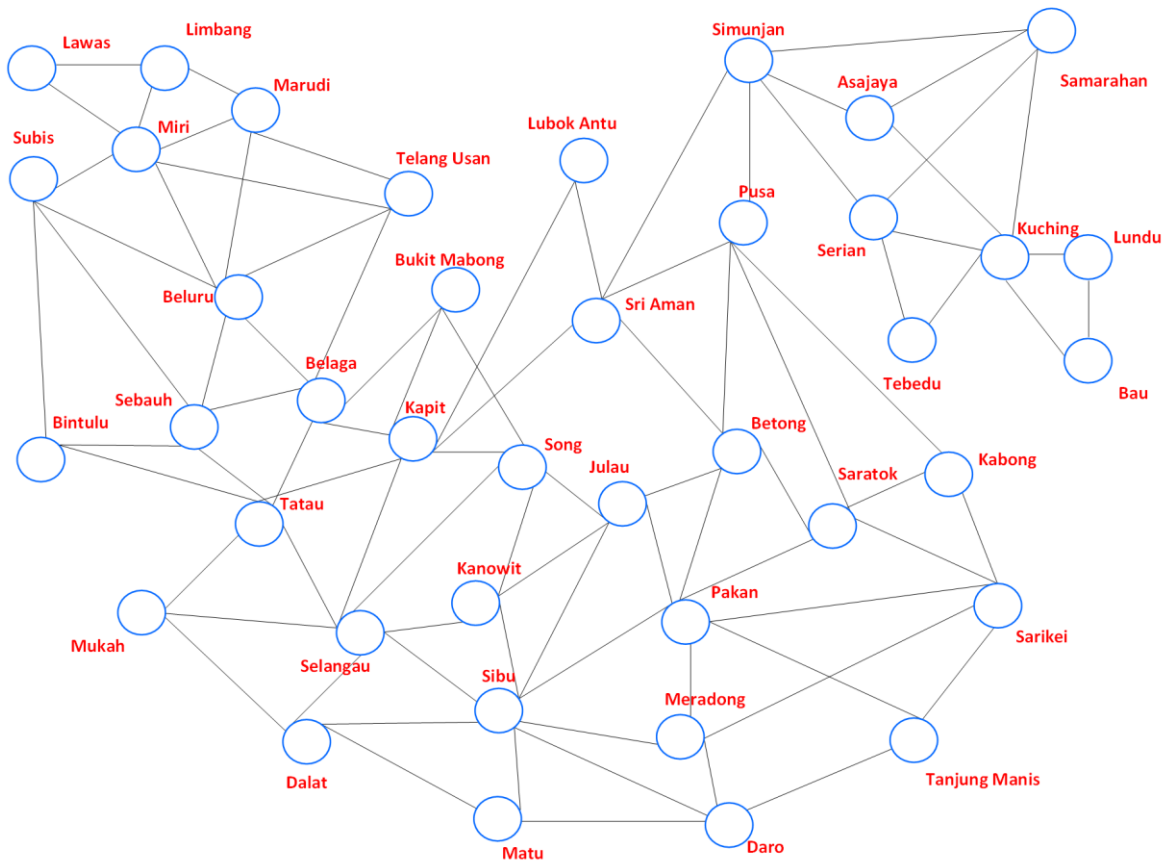


Figure 5.4: Plot of weight matrix of district in Sarawak

Figure 5.4 depicts the weight matrix for these 40 districts. Within Figure 5.4, the blue dots symbolize a value of 1, while empty cells signify a value of 0.

Table 5.2: Incidence rate of districts in Sarawak for the year and all three waves

No.	Districts	Incidence 2021	Incidence W1	Incidence W2	Incidence W3
1	Asajaya	81.622	12.438	67.071	2.112
2	Bau	142.355	22.909	113.956	5.489
3	Belaga	95.680	16.798	75.815	3.066
4	Beluru	43.631	27.426	12.092	4.112
5	Betong	125.058	24.515	89.414	11.128
6	Bintulu	104.049	47.680	48.167	8.200
7	Bukit Mabong	122.796	91.482	25.603	5.711
8	Dalat	55.468	23.833	18.300	13.335
9	Daro	21.461	2.515	12.270	6.674
10	Julau	76.827	46.631	26.022	4.174
11	Kabong	28.417	12.605	11.953	3.857

Table 5.2 continued

12	Kanowit	118.987	67.287	41.781	9.919
13	Kapit	150.513	100.027	44.934	5.550
14	Kuching	122.205	26.002	85.165	11.037
15	Lundu	115.624	21.446	89.190	4.988
16	Marudi	28.930	9.395	14.545	4.989
17	Matu	35.731	15.751	14.832	5.148
18	Meradong	127.838	86.605	34.829	6.404
19	Miri	81.449	32.051	32.714	16.682
20	Mukah	115.765	48.633	56.700	10.431
21	Pakan	117.578	53.421	60.794	3.363
23	Pusa	127.013	13.294	104.412	9.306
24	Samarahan	79.207	16.377	59.781	3.047
25	Saratok	133.044	39.169	83.704	10.170
26	Sarikei	93.598	37.375	48.570	7.652
28	Sebauh	29.960	18.171	10.639	1.148
29	Selangau	119.229	69.995	46.008	3.225
30	Serian	158.076	34.565	113.644	9.865
31	Sibu	112.519	54.054	48.640	9.824
32	Simunjan	85.913	15.105	65.919	4.888
33	Song	255.998	126.995	119.967	9.035
34	Sri Aman	102.975	36.447	62.575	3.951
36	Subis	73.364	33.758	36.010	3.595
37	Tanjung Manis	72.111	44.047	14.976	13.088
38	Tatau	74.028	37.500	35.338	1.190
39	Tebedu	77.045	18.310	56.634	2.100
40	Telang Usan	39.411	29.644	6.836	2.930

The incidence rate for both the state and district is derived by combining their data.

The incidence rate for the state has already been calculated and is detailed in Section 5.1.1.

Therefore, the incidence rate for Sarawak's districts is subsequently calculated.

Table 5.2 displays the COVID-19 incidence rates in Sarawak for 2021 and three waves, showing significant variations among districts. In 2021, the district with the highest incidence rate was Song (255.99), followed by Serian (158.07), Kapit (150.51), and Marudi (28.93). On the other end of the spectrum, the districts with the lowest incidence rates were Telang Usan (39.41) and Sebauh (29.96).

Furthermore, individual waves' highest and lowest incidence rates were observed in different districts. For example, in wave 1, Song (126.99) and Daro (2.51) had the highest

and lowest incidence rates, respectively. In wave 2, Song (119.96) and Telang Usan (6.83) exhibited the highest and lowest rates, and in wave 3, Dalat (13.33) and Sebauh (1.14) had the highest and lowest incidence rates. Based on the data in Table 5.2, the overall incidence rate of districts was highest during the second wave of the pandemic, followed by the first wave and then the third wave.

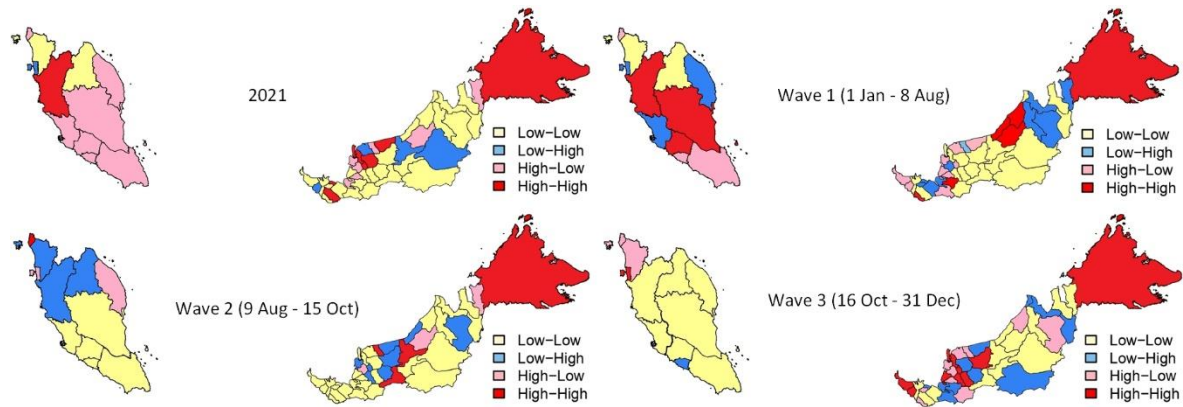


Figure 5.5: LISA diagram of two-level of Malaysia

The LISA map (Figure 5.5) depicts spatial clusters of incidence rates in Malaysia with a two-level analysis for 2021. Specifically, the high-high clusters, represented in red, are primarily concentrated in Kedah and Sabah in Malaysia. In addition, these high-high clusters in the Sarawak district encompass areas such as Mukah, Daro, Meradong, Sibu, Asajaya, and Serian.

In the first wave, it is evident that the west-south region of Sarawak state features high-low (pink) and low-high (blue) clusters. Notably, in the second wave, these clusters have shifted to the state's eastern side. Finally, in the third wave, they emerged in the south-eastern and north-eastern regions of the state of Malaysia.

Additionally, it is essential to consider that COVID-19 transmission can be influenced by travel patterns, which can, in turn, impact the incidence rate. As a result,

another type of analysis, specifically the state-to-district analysis, has been conducted in addition to the two-level analysis.

5.1.3 LISA for State to District

The discussion on the weight of states and districts has been previously covered in Sections 5.1.1 and 5.1.2. Therefore, the focus now shifts to the computation of the submatrix for the state-to-district weight matrix, as outlined in Equation 5.3.

Malaysia boasts two prominent domestic airlines, AirAsia and Malaysian Airlines (Chaudhuri & Srivastava, 2022; Montes-Orozco et al., 2023). This study adheres to the official domestic flight routes provided by these two airlines, as documented in their sources (AirAsia, 2023; AirAsia at Kuala Lumpur Airport, 2023; Malaysia Airlines, 2023). A connectivity matrix is generated from these routes to establish the connectivity patterns, as elaborated below.

$$\text{State to District Weight Matrix} = \begin{bmatrix} \text{State}_{16 \times 16} & \text{District to State}_{16 \times 40} \\ \text{State to District}_{40 \times 16} & \text{Districts}_{40 \times 40} \end{bmatrix}$$

Equation 5.3

Table 5.3: Connectivity matrix of AirAsia domestic flight route

State	City	Kedah		Pulau Pinan g	Kuala Lump ur	Kelanta n	Terengga nu	Johor	District of Sarawak State				Sabah			
		Langk awi	Alor setar	Penan g	Kuala Lump ur	Kota Bharu	Kuala Terengga nu	Johor Bahru	Kuchi ng	Sibu	Bintul u	Miri	Labua n	Kota Kinabal u	Sandak an	Tawa u
Kedah	Langkawi				1											
	Alor setar				1											
Pulau Pinang	Penang				1			1	1					1		
Kuala Lumpur	Kuala Lumpur	1	1	1		1	1	1	1	1	1	1	1	1	1	1
Kelantan	Kota Bharu				1											
Terengga nu	Kuala Terengga nu				1											
Johor	Johor Bahru			1	1				1	1		1		1	1	1
District of Sarawak State	Kuching			1	1			1		1	1	1		1		
	Sibu				1			1	1							
	Bintulu				1				1							
	Miri				1			1	1					1		
Sabah	Labuan				1											
	Kota Kinabalu			1	1			1	1			1			1	1
	Sandakan				1			1						1		
	Tawau				1			1						1		
							Represents the Districts									

Table 5.4: Connectivity matrix of Malaysian Airlines domestic flight route

State	City	Kedah		Pulau Pinan g	Kuala Lump ur	Kelanta n	Terengga nu	Johor	District of Sarawak State				Sabah			
		Langk awi	Alor setar	Penan g	Kuala Lump ur	Kota Bharu	Kuala Terengga nu	Johor Bahru	Kuchi ng	Sibu	Bintul u	Miri	Labua n	Kota Kinabal u	Sandak an	Tawa u
Kedah	Langkawi					1										
	Alor setar					1										
Pulau Pinang	Penang					1										
Kuala Lumpur	Kuala Lumpur					1										
Kelantan	Kota Bharu	1	1	1	1		1	1	1	1	1	1	1	1	1	1
Terengga nu	Kuala Terengga nu					1										
Johor	Johor Bahru					1										
District of Sarawak State	Kuching					1				1						
	Sibu					1			1		1	1	1		1	
	Bintulu					1				1						
	Miri					1				1						
Sabah	Labuan					1				1					1	
	Kota Kinabalu					1									1	

Sandakan					1				1			1	1		1
Tawau					1									1	
						Represents the Districts									

Subsequently, merge these connectivity matrices (as represented in Tables 5.3 and 5.4) to depict the linkages between states and districts, as visualized in Table 5.5. In scenarios where a state lacks any flights to a particular district, the value in the corresponding cell of the connectivity matrix is treated as zero. It implies that the vacant cells in Table 5.3 to 5.5 indicate zero values.

Table 5.5: Connectivity matrix of state to district using both domestic flight route

State/District	Kuching	Sibu	Bintulu	Miri
Penang	●			
Kuala Lumpur	●	●	●	●
Johor	●	●		●
Sabah	●			●

Next, update the values in the submatrices that represent domestic flight routes between states. Then, integrate these revised values into Equation 5.3 to create the weight matrix for states to districts.

To calculate the state's incidence rate to the district, this study combine the information from the state of Malaysia and the district of Sarawak, as previously computed and discussed in the earlier sections (Sections 5.1.1 and 5.1.2). With these data, the focus now shifts towards generating a LISA map based on the Local Moran's I value.

The LISA map (depicted in Figure 5.6) offers insights into the spatial clustering of incidence rates in Malaysia, employing a state-to-district analysis specifically for the year 2021. In this analysis, the high-high clusters, indicated in red, are notably concentrated in Pulau Pinang, Perak, and Johor within Peninsular Malaysia. Additionally, these high-high

clusters also encompass areas within the Sarawak district, including Mukah, Daro, Meradong, Sibu, Asajaya, and Serian.

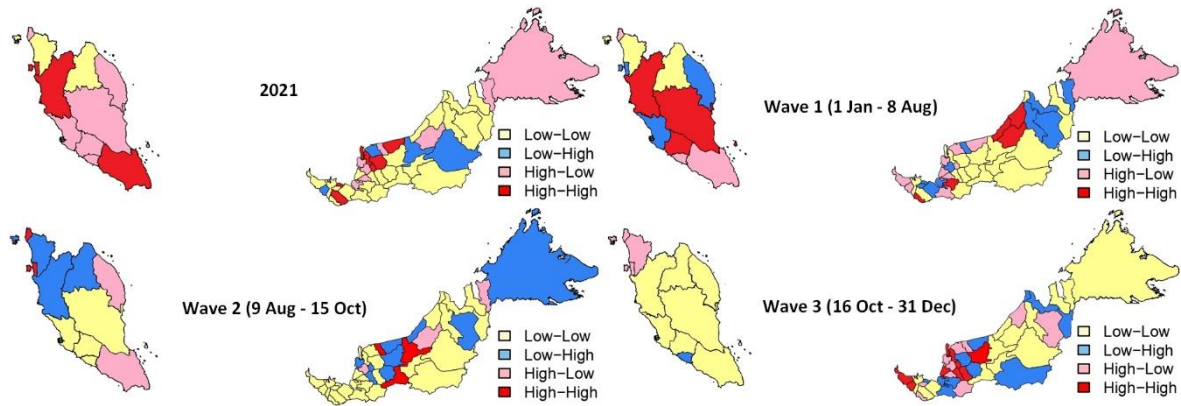


Figure 5.6: LISA diagram of State to District of Malaysia

Moreover, it is noticeable that in wave 1, there is a higher number of areas falling into the High-High and High-Low clusters, while in wave 2, a more significant number of areas are classified as Low-High or Low-Low, particularly when compared to wave 3.

5.2 Global Moran's I value

The significance of the Global Moran's I value in the context of COVID-19 incidence rates lies in its ability to measure spatial autocorrelation (Kianfar & Mesgari, 2022). Spatial autocorrelation is a statistical metric that characterizes the extent to which the values of a variable at one location are related to the values of the same variable at nearby locations. When applied to incidence rates, a positive Global Moran's I value indicates a tendency for these rates to cluster together in space. This implies that areas with high incidence rates are more likely to have nearby areas with high incidence rates as well.

Table 5.6 provides Moran's I statistics and associated *p*-values for incidence rates across various time periods, computed using first-order queen contiguity spatial weights. These statistics evaluate spatial dependence, helping to determine whether incidence rates

exhibit spatial clustering (positive spatial autocorrelation) and whether these clusters have statistical significance.

Table 5.6: Global Moran's I values of district, state, two-level and state to district of 2021 and different wave of 2021

Spatial dependence		Incidence rate			
		2021	W1	W2	W3
Districts	Moran's I statistic	0.213	0.387	0.339	0.273
	<i>p</i> -value	0.022	7.339e-05	0.0004	0.004
States	Moran's I statistic	0.080	0.585	0.242	0.199
	<i>p</i> -value	0.527	0.005	0.183	0.251
Two-Level	Moran's I statistic	0.235	0.429	0.422	0.376
	<i>p</i> -value	0.014	1.54e-05	2.09e-05	0.0001
States to Districts	Moran's I statistic	0.228	0.393	0.380	0.375
	<i>p</i> -value	0.010	1.907e-05	3.393e-05	4.25e-05

For districts, two-level comparisons, and states to districts, the Moran's I statistics consistently yield positive values, signifying a propensity for incidence rates to cluster. These statistics are statistically significant, as indicated by *p*-values below 0.05, indicating that the clustering is not occurring randomly. However, in the case of states, Moran's I is only significant in wave 1, where the *p*-value falls below 0.05. It suggests that during wave 1, the incidence rates at the state level exhibit notable spatial autocorrelation.

In 2021 and Wave 3, the incidence rates for the state level display the lowest significant spatial autocorrelation, with *p*-values exceeding 0.05. It implies that the spatial clustering patterns in those periods at the state level are insignificant. Furthermore, the global spatial autocorrelation at the state level in 2021 is weaker than that at the district level in "Sarawak Districts." This discrepancy could be attributed to our contiguity-based modeling needs to fully capture the connectivity of Kuala Lumpur, Malaysia's capital, with other states and federal territories. Kuala Lumpur only shares contiguity with Selangor, which needs to

accurately reflect the real connectivity it should have with other states and federal territories in Malaysia.

The findings presented in Table 5.3 indicate that incorporating state-level entities into the analysis of Sarawak districts results in higher global Moran's I statistics for the year 2021 and all three waves. It suggests that more than focusing on Sarawak districts is needed to consider state-level spatial relationships for comprehending the spatial distribution of disease incidence in Sarawak.

The calculated p -values for the "States to Districts" comparison are consistently below 0.05, indicating the statistical significance of spatial clustering in incidence rates at this level. In other words, the distribution of incidence rates among districts is not random; it exhibits meaningful spatial patterns and tends to be more similar to neighboring districts.

When comparing the "States to Districts" analysis with the district-level analysis, it becomes evident that including state-level information strengthens the spatial autocorrelation patterns. It suggests that the incidence rates at the district level are influenced by their immediate district neighbors and neighboring states. Therefore, considering state-level spatial associations is crucial for comprehensively understanding disease dynamics in Sarawak.

In summary, the "States to Districts" analysis provides compelling evidence of significant spatial clustering in disease incidence rates at the district level when considering state-level associations. It emphasizes the importance of considering spatial relationships at district and state levels to gain a holistic insight into disease transmission and prevalence patterns in Sarawak.

Additionally, an examination was conducted at the district-to-state level, which revealed negative values for the year 2021 (-0.026) and wave 1 (-0.049), but they became

positive in waves 2 (0.078) and wave 3 (0.129). Notably, for all four scenarios, the p -values exceeded 0.05 (0.939, 0.758, 0.3418, and 0.149).

The overall analysis of Table 5.3 highlights spatial dependence across different levels, including state, district, two-level, and state-to-district comparisons, providing valuable insights into the spatial patterns of disease incidence rates in Sarawak. Moran's I statistics consistently reveal positive values for all levels except for states in wave 1, indicating a tendency for spatial clustering (positive spatial autocorrelation). Importantly, all significant p -values (< 0.05) for districts, two-level, and state-to-district comparisons underscore the non-random nature of the spatial patterns, implying that neighboring areas share similarities in incidence rates. The inclusion of state-level associations reinforces these patterns, underscoring the significance of considering both state and district-level spatial relationships for a comprehensive understanding of disease evolution in Sarawak.

5.3 Centrality of Weight Matrix

In order to determine which states and districts have the highest and lowest impact on each other, this study analyzed various centrality measures of nodes in the weight matrix (Dou et al., 2021; Chaudhuri & Srivastava, 2022). The following steps were taken to identify the impact (Salavati et al., 2019):

- a) **Prepare the data:** A supra-adjacency matrix of state to state, state to district, and district to district.
- b) **Calculate the centrality measures of the nodes of the weight matrix:** This study used R to calculate the centrality measures using the `igraph` package and `graph`, `degree`, `betweenness`, and `eigen_centrality` command. This study considered these centrality measures:

- **Degree centrality:** The degree centrality of a node is the number of other nodes that it is connected to. Nodes with high degree centrality are more likely to spread the virus because they have more contacts with other districts (Lv et al., 2019).
 - **Betweenness centrality:** The betweenness centrality of a node is the number of times that it falls on the shortest path between two other nodes. Nodes with high betweenness centrality are important because they control the flow of information through the network (Gómez, 2019).
 - **Eigenvector centrality:** The eigenvector centrality of a node is a measure of its influence in the network. Nodes with high eigenvector centrality are important because they are more likely to be connected to other important nodes (Gómez, 2019).
- c) **Identify the states with the highest centrality measures:** Once centrality measures have been calculated for all of the nodes in the supra-adjacency matrix, the analysis can identify the states or districts with the highest centrality measures. These states or districts are the ones that are most likely to spread the virus to other states or districts.
- d) **Examine the supra-adjacency matrix to identify the districts that are connected to these states:** Once the states or districts with the highest centrality measures have been identified, the supra-adjacency matrix is examined to identify the states or districts connected to them. These districts are at the highest risk of infection from these states.
- e) **Interpret the results:** The study identified the states likely to spread COVID-19 cases to other districts and districts with a significant risk of infection from these states based on the analysis's results.

5.3.1 Centrality of State to District (Flight Route)

The three sub-figures within Figure 5.7 illustrate the interconnections among Malaysian states and the Sarawak district via flight routes, as represented in Table 5.5. These flight routes encompass direct connections provided by AirAsia and Malaysian Airlines. All three of these centrality metrics can be utilized to pinpoint areas with a heightened risk of COVID-19 transmission.

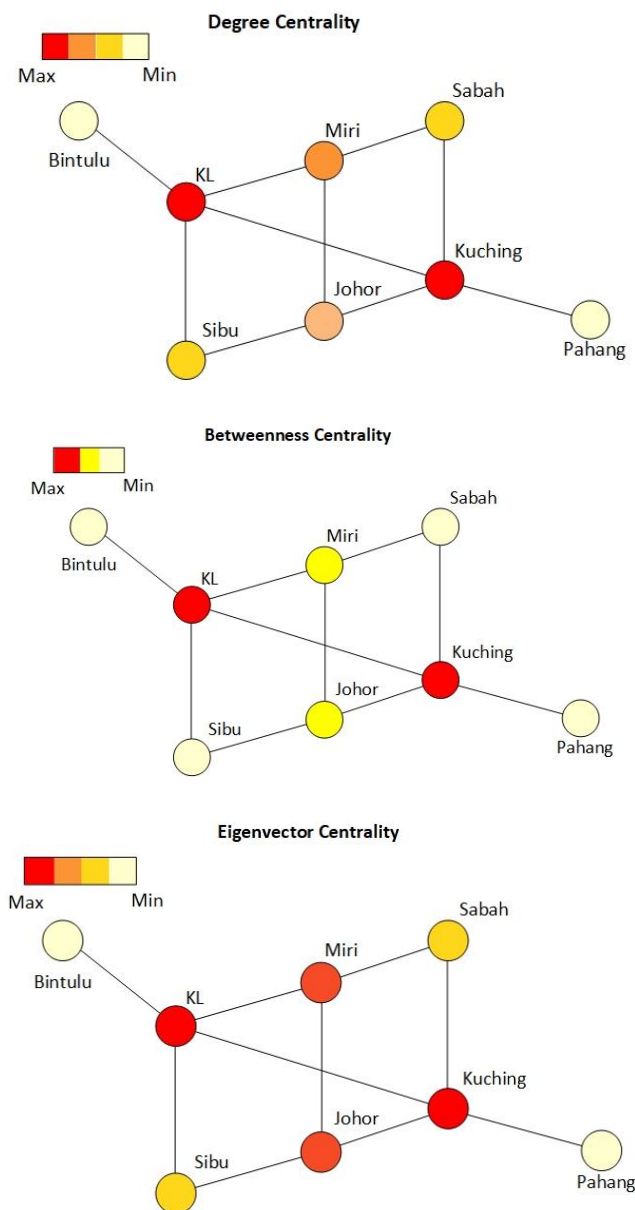


Figure 5.7: Centrality between states of Malaysia and the district of Sarawak according to flight route

Regions displaying a substantial degree of centrality are at greater risk due to their extensive connections with other areas, thus increasing their potential exposure to the virus. Similarly, areas with elevated betweenness centrality are also more susceptible because they lie on the shortest paths between many other regions, making them more likely to be involved in virus transmission. Furthermore, areas with high eigenvector centrality are also at an elevated risk, as they are linked to other critical places within the network, potentially facilitating virus spread.

Regarding degree centrality, the cities of Kuala Lumpur and Kuching display the most significant degrees, with Miri and Johor following closely. Kuala Lumpur is Malaysia's capital, while Kuching is the capital of the Sarawak district. Kuala Lumpur and Kuching have the shortest connections to other locations, resulting in a notably high betweenness degree compared to other areas. Similarly, when considering Eigenvector centrality, Kuala Lumpur and Kuching top the rankings, with Miri and Johor as the next in line. All three sub-figures consistently affirm that Kuala Lumpur and Kuching hold the highest centrality values. It underscores that these two cities are particularly vulnerable to the transmission of COVID-19 within Malaysia. The increased risk is due to Kuala Lumpur and Kuching occupying pivotal positions within the Malaysian air transportation network, establishing them as primary hubs for people traveling to and from various destinations.

5.3.2 Centrality of State to State

The impact of COVID-19 varies across different states in Malaysia. Therefore, using a weight matrix becomes essential for elucidating the spatial relationships among these states (Dong & Guo, 2021). This study aimed to uncover the centrality relationships among these states based on the weight matrix (as depicted in Figure 5.8).

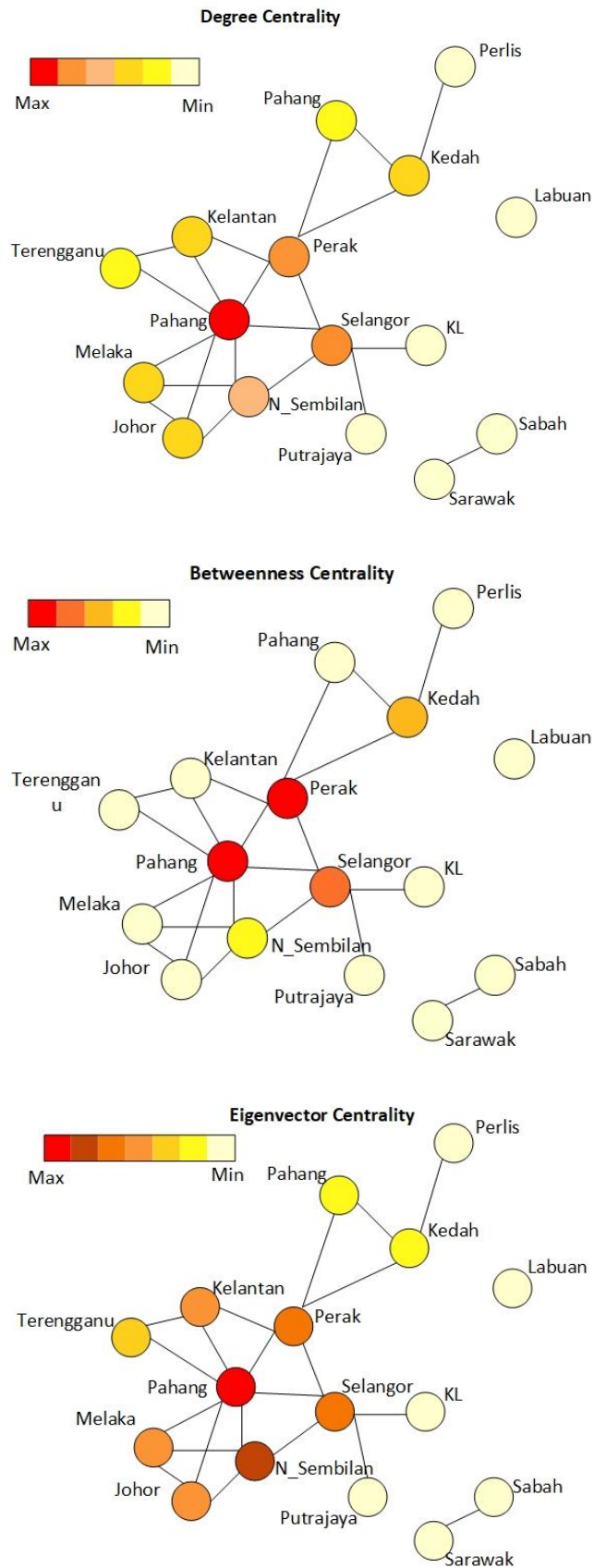


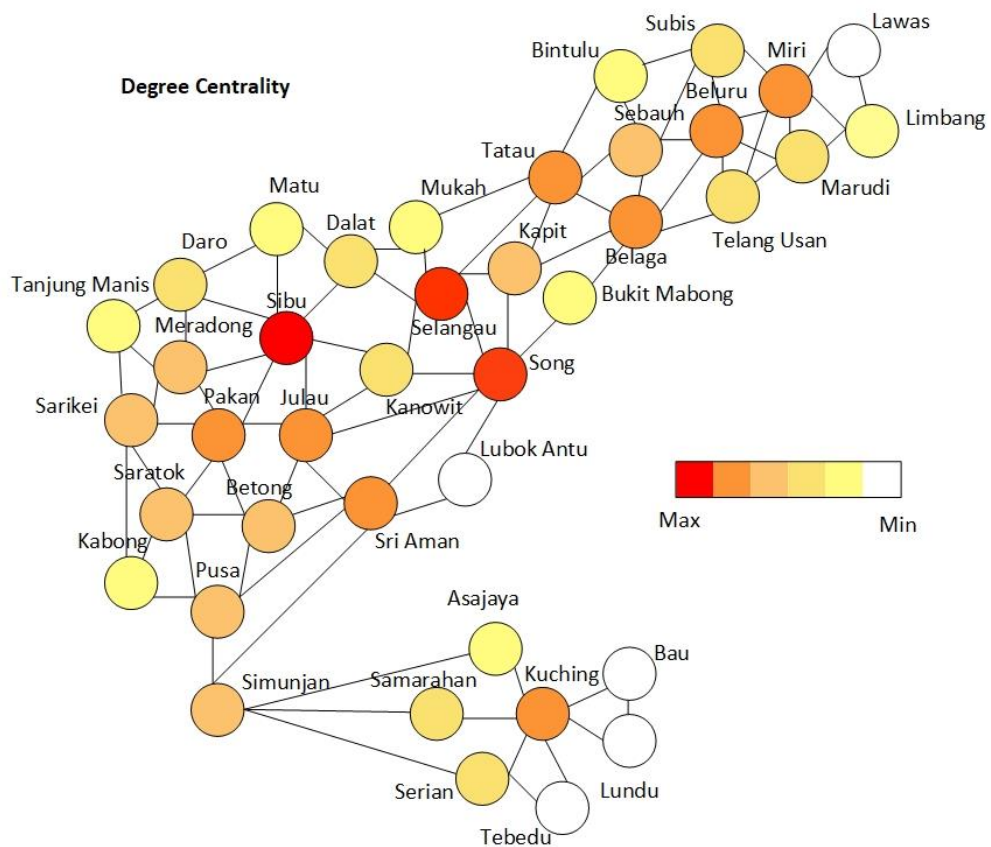
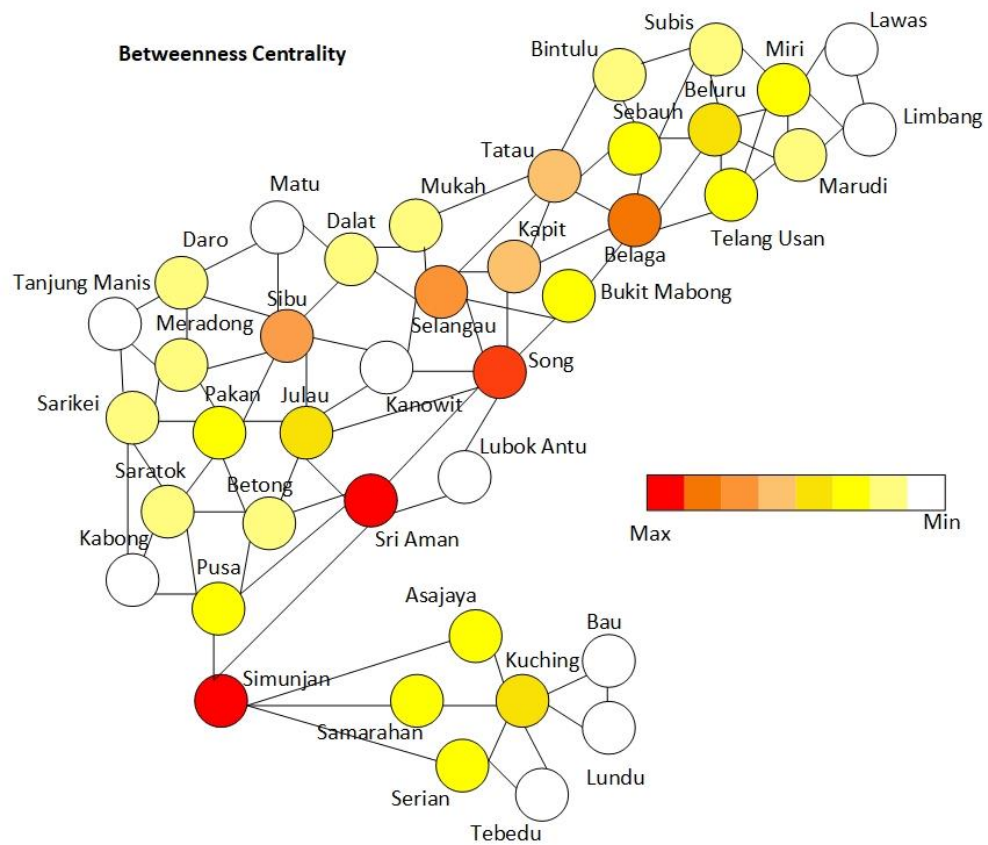
Figure 5.8: Centrality between states to state of Malaysia

The sub-figures within Figure 5.8 reveal that Pahang exhibits the highest centrality degree, followed by Perak and Selangor. It indicates that these states have strong connections to the other states, making them susceptible to COVID-19 transmission. Moreover, Pahang, and Perak, have the shortest connections to most other states. As a result, their betweenness centrality is notably higher than the remaining states. Additionally, Pahang followed by Perak and Selangor, displays higher Eigenvector centrality, signifying their connections to other pivotal states. In conclusion, Pahang emerges as the state with the highest centrality in all three categories. Consequently, it can be inferred that Pahang carries a higher risk of COVID-19 transmission.

5.3.3 Centrality of District to District (Sarawak)

As previously mentioned, centrality measures play a crucial role in identifying the key node within a network. Therefore, to determine the pivotal district in Sarawak state concerning the dissemination of COVID-19, this study applied centrality analysis to the district's weighted matrix (as depicted in Figure 5.9).

Figure 5.9 shows that Simunjan and Sri Aman have the highest centrality scores, with Song closely following. Similarly, Sibü has the highest betweenness centrality, followed by Selangau and Song. In terms of eigenvector centrality, Sibü, Selangau, and Song lead the list, with Pakan and Julau slightly behind. Overall, Sibü's top rankings in degree, betweenness, and eigenvector centrality highlight its crucial role in Sarawak's network. Additionally, Song ranks in the top three across all categories. Thus, Sibü and Song emerge as the most critical districts in the network concerning the spread of COVID-19.



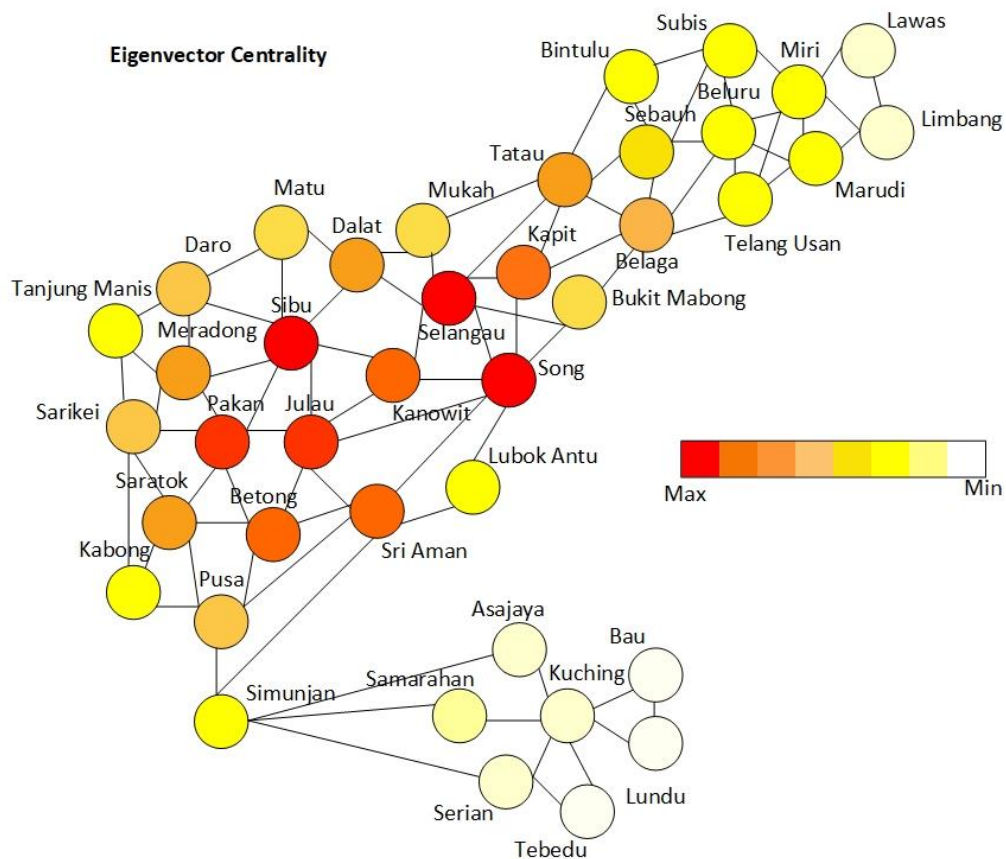


Figure 5.9: Centrality between districts to district of Sarawak state

5.4 Summary

The primary goal of this chapter was to identify spatial clusters of high and low COVID-19 incidence rates within Malaysia, focusing on the state of Malaysia and the district of Sarawak, and examining the impact of travel on virus spread between these areas. Moran's I values were utilized, along with a detailed examination of the weight matrix for state and two-level analyses. Both Local and Global Moran's I values were considered, with incidence rates calculated as new cases per 1,000 people.

In 2021, Selangor had the highest incidence rate, followed by Kuala Lumpur and Negeri Sembilan, while the lowest rates were in Perlis, Perak, and Sabah. Labuan had the highest incidence in Wave 1 and the lowest in Wave 2, whereas Sarawak saw the highest incidence in Wave 2, indicating a significant surge in cases during this period. The study

calculated Local Moran's I values based on incidence rate and weight, leading to the creation of a LISA map. Using an existing weight matrix for Malaysia, a new matrix for two-level analysis, combining the state of Malaysia and the Sarawak district, was generated with the supra adjacency method to account for state-to-district connections.

The 2021 LISA map showed high COVID-19 incidence rates in various state of Malaysia, identifying high-high clusters in Perlis, Terengganu, Melaka, and Johor throughout the year. Wave-specific data highlighted Perak and Sabah in high-high clusters during wave 1, Kedah, Terengganu, and Sarawak in wave 2, and Kelantan in wave 3. Pahang, Selangor, and Putrajaya were in a high-low cluster for the year, while Perak and Sarawak were in a low-high cluster overall, with variations in specific waves.

In Sarawak, the two-level analysis map of 2021 revealed significant differences in incidence rates among districts. Song consistently had the highest rates, followed by Serian, Kapit, and Marudi, while Telang Usan and Sebauh had the lowest. The map also identified high-high clusters primarily in Kedah and Sabah, and in Sarawak's districts like Mukah, Daro, Meradong, Sibul, Asajaya, and Serian. Wave-specific analysis showed clustering shifts across Sarawak, from the west-south in wave 1 to the eastern and southeastern-northeastern regions in waves 2 and 3. State-to-district connections were examined using the supra adjacency method, focusing on flight routes from Malaysian Airlines and Air Asia as significant factors in virus spread. The LISA map analysis of these connections revealed high-high clusters in Pulau Pinang, Perak, and Johor, and in several districts of Sarawak, particularly in wave 1.

Global Moran's I values were calculated alongside Local Moran's I to assess spatial dependence across various levels, such as state, district, two-level, and state-to-district comparisons. The analysis consistently showed positive Moran's I values, indicating spatial

clustering except for states in wave 1. Significant p -values (< 0.05) for districts, two-level, and state-to-district comparisons confirmed the non-random nature of these spatial patterns, suggesting similarities in neighboring areas' incidence rates. This comprehensive approach, considering both state and district-level spatial relationships, provided valuable insights into the disease evolution in Sarawak, reinforcing the observed spatial patterns.

Centrality analysis for the state of Malaysia, the district of Sarawak, and state-to-district connections identified key nodes in COVID-19 transmission. Analyzing flight routes revealed that Kuala Lumpur and Kuching had the highest centrality scores, followed by Miri and Johor. Kuala Lumpur and Kuching had the shortest connections to other locations, resulting in high betweenness centrality. In terms of eigenvector centrality, Kuala Lumpur and Kuching also ranked highest, emphasizing their key roles in Malaysia's air transportation network and vulnerability to COVID-19 transmission. Among Malaysian states, Pahang had the highest degree centrality, followed by Perak and Selangor, indicating their strong connections to other states and susceptibility to virus transmission. Pahang and Perak had the shortest connections to most other states, resulting in higher betweenness centrality. Pahang, Perak, and Selangor also had higher eigenvector centrality, highlighting their connections to other pivotal states.

In Sarawak, centrality analysis among districts showed Simunjan and Sri Aman with the highest centrality scores, closely followed by Song. Sibu had the highest betweenness centrality, followed by Selangau and Song. In terms of eigenvector centrality, Sibu, Selangau, and Song were at the forefront. Sibu's prominent rankings in all centrality measures underscored its critical role in Sarawak's network, with Song also emerging as a significant district in the spread of COVID-19.

CHAPTER 6

CONCLUSION AND RECOMMENDATION

This study comprehensively analyzed the spatiotemporal variation of district-level COVID-19 incidence rates and their real relationship scenario on sociodemographic factors in Sarawak, Malaysia. Furthermore, the analysis was conducted using spatial autocorrelation analysis and regression models, which included the global non-spatial linear, global spatial lag, and spatial error models. The study's results demonstrated that the spatial regression model performed better than the non-spatial model. Moreover, the SLM displayed notable effectiveness within the specific spatial model. Additionally, the third wave exhibited the weakest correlation across all waves compared to the first and second waves.

This study investigated the spatial-temporal dynamics of COVID-19 at the district level in Sarawak, Malaysia. Utilizing spatial panel data modeling, spatial autocorrelation analysis, and spatial regression models, the research examined the spatial dependence and identified influential sociodemographic factors associated with COVID-19 transmission. The analysis focused on the spatial distribution of confirmed COVID-19 cases across the 40 districts of Sarawak from 1 January to 31 December 2021 and incorporated district-wise demographic data and various socioeconomic indicators. The findings highlighted the significance of socioeconomic disparities and infrastructure quality in affecting the distribution and intensity of COVID-19 cases.

The study emphasized the importance of considering sociodemographic indicators such as population growth rate, non-citizen population, household income, and basic amenities in understanding the impact and trajectory of COVID-19. The results suggested that targeted public health interventions and resource allocations, informed by spatial

dependencies and demographic factors, could effectively mitigate the impact of pandemics in diverse districts. The analysis, conducted using R-studio and incorporating LISA for spatial autocorrelation, underscored the importance of integrating spatial analysis in public health strategies to address regional health disparities.

Furthermore, this research focused on analyzing the spatial autocorrelation in Malaysia at both the state and district levels, with a specific focus on Sarawak. The study examined the impact of flight routes on the transmission of COVID-19 between states and districts, using supra-adjacency matrices to calculate both local and global Moran's I values. Additionally, centrality measures were used to identify key nodes in districts, states, and the connections between them. This analysis provided insight into the spatial patterns of disease incidence rates in Sarawak, highlighting spatial dependence across various levels. Moran's I statistics consistently showed positive values, indicating spatial clustering. Notably, all significant p -values (< 0.05) for districts, two-level, and state-to-district comparisons pointed to the non-random nature of the spatial patterns, suggesting that neighboring areas shared similarities in incidence rates. Kuala Lumpur and Kuching had the highest significance in the connection from states to districts, while Pahang demonstrated the greatest significance at the state level, and Sibu and Sri Aman showed the most significance at the district level.

6.1 Contribution

The following objectives outline the main contribution of this research.

Objective 1: To describe the evolution of Sarawak COVID-19 data in different waves of pandemic.

The methodology of spatial autocorrelation of Sarawak COVID-19 data involves three steps. The first step is to divide the period into three periods/waves using change point analysis. Then, the incidence rates are calculated. Lastly, normality is checked.

Objective 2: To quantify the spatial dependence of district-wise in Sarawak and state-wise in Malaysia aggregated COVID-19 confirmed cases using spatial panel data modelling.

The methodology for quantifying spatial dependence in district-wise COVID-19 confirmed cases involved several steps:

- a. A spatial weight matrix was constructed to capture the spatial relationships between districts. This weight matrix was based on the geographical contiguity between districts and represented their spatial connectivity.
- b. The Global Moran's I statistic was computed to assess the overall spatial autocorrelation in the COVID-19 cases across all districts. The Global Moran's I value indicates whether the pattern of COVID-19 cases is clustered, dispersed, or randomly distributed across Sarawak and Malaysia.
- c. LISA analysis was performed to identify local clusters of districts and state with high or low COVID-19 incidence. LISA provides insights into which districts or state have significant spatial associations with their neighboring districts, i.e., whether they belong to clusters of high-high, low-low, high-low, or low-high COVID-19 cases.
- d. Moran scatter plots were generated to visualize the spatial autocorrelation patterns and identify potential outliers or influential districts and state in the data.

These spatial panel, data modeling techniques, provide a comprehensive understanding of the spatial relationships and dependencies among district-wise and and

state-wise COVID-19 cases and contribute to a more nuanced analysis of the pandemic's spatial dynamics in Sarawak and Malaysia.

Objective 3: To construct spatial regression models for finding the spatial relationship between district-wise COVID-19 cases and their demographical, social, and economic factors.

The regression model is divided into spatial and non-spatial categories in this study. OLS are under the non-spatial model SLM, and SEM is under the spatial models category. Whereas all were used to predict the relationship between district-wise COVID-19 cases and its demographical, social, and economic factors.

6.2 Limitations

The primary limitation in the study stems from the utilization of first-order contiguity in the spatial analysis. While this approach helps capture immediate spatial relationships, it might overlook the influence of more distant neighboring regions that could spatially impact the spatial distribution of COVID-19 cases during Wave 3 in Sarawak. This limitation could have led to an underestimation of the interconnectedness between various regions, thereby affecting the accuracy of the findings. The exclusion of higher-order contiguity in the spatial analysis could have limited the understanding of the broader spatial dynamics within the region during Wave 3. As a result, the study may need to fully capture the intricate spatial interdependencies between neighboring regions, potentially leading to a less comprehensive analysis of the factors influencing the spread of COVID-19.

6.3 Future Works

In order to address the limitations associated with using first-order contiguity in spatial analysis, future work should explore the incorporation of higher-order contiguity into the analytical framework. By including higher-order contiguity in the spatial analysis, the

study can gain a more comprehensive understanding of the extended spatial relationships and interdependencies between regions, providing a more accurate portrayal of the spatial dynamics of COVID-19 in Sarawak. Furthermore, conducting a comparative analysis between the results obtained from using first-order contiguity and those from including higher-order contiguity can provide valuable insights into the importance of considering various degrees of spatial relationships in understanding the spread of COVID-19. This comparative approach can contribute to refining spatial analysis methodologies and enhance the overall robustness of the study's findings, specifically in the context of Wave 3 in Sarawak.

REFERENCES

- Adekunle, I. A., Onanuga, A. T., Akinola, O. O., & Ogunbanjo, O. W. (2020). Modelling spatial variations of coronavirus disease (COVID-19) in Africa. *Science of the Total Environment*, 729, 1-4.
- Ahmar, A. S., & Boj, E. (2020). Will COVID-19 confirmed cases in the USA reach 3 million? A forecasting approach by using Sutte ARIMA Method. *Current Research in Behavioral Sciences*, 1, 100002.
- Al-Kindi, K. M., Alkharusi, A., Alshukaili, D., Al Nasiri, N., Al-Awadhi, T., Charabi, Y., & El Kenawy, A. M. (2020). Spatiotemporal assessment of COVID-19 spread over Oman using GIS techniques. *Earth Systems Environmental Challenges*, 4, 797-811.
- Al Awaidy, S. T., Khamis, F., Al Rashidi, B., Al Wahaibi, A. H., Albahri, A., & Mahomed, O. (2021). Epidemiological characteristics of 69,382 COVID-19 patients in Oman. *Journal of Epidemiology and Global Health*, 11, 326-337.
- Alexakis, C., Eleftheriou, K., & Patsoulis, P. (2021). COVID-19 containment measures and stock market returns: An international spatial econometrics investigation. *Journal of Behavioral Experimental Finance*, 29, 1-4.
- Alkhamis, M. A., Al Youha, S., Khajah, M. M., Haider, N. B., Alhardan, S., Nabeel, A., Al Mazeedi, S., Al-Sabah, S. K. (2020). Spatiotemporal dynamics of the COVID-19 pandemic in the State of Kuwait. *International Journal of Infectious Diseases*, 98, 153-160.
- AlQadi, H., Bani-Yaghoub, M., Balakumar, S., Wu, S., & Francisco, A. (2021). Assessment of Retrospective COVID-19 Spatial Clusters with Respect to Demographic Factors:

- Case Study of Kansas City, Missouri, United States. *International Journal of Environmental Research Public Health*, 18(21), 11496.
- Aw, S. B., Teh, B. T., Ling, G. H. T., Leng, P. C., Chan, W. H., & Ahmad, M. H. (2021). The COVID-19 pandemic situation in Malaysia: lessons learned from the perspective of population density. *International Journal of Environmental Research and Public Health*, 18(12), 1-21.
- Barreal, J., & Jannes, G. (2023). Spatial Dependencies and Neighbour Interactions of Wildfire Patterns in Galician Mountain Areas (NW Spain). *Fire*, 6(4), 165.
- Ben Abacha, A., Sarrouiti, M., Demner-Fushman, D., Hasan, S. A., & Müller, H. (2021). Overview of the vqa-med task at imageclef 2021: Visual question answering and generation in the medical domain. Proceedings of the CLEF 2021 Conference and Labs of the Evaluation Forum-working notes, Bucharest, Romania, 21-24 September 2021.
- Bilal, M., Khan, K. I. A., Thaheem, M. J., & Nasir, A. R. (2020). Current state and barriers to the circular economy in the building sector: Towards a mitigation framework. *Journal of Cleaner Production*, 276, 123250.
- Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., & Blanchet, G. (2009). Package ‘spdep’. The Comprehensive R Archive Network.
- Briz-Redón, Á., Iftimi, A., Correcher, J. F., De Andrés, J., Lozano, M., & Romero-García, C. (2022). A comparison of multiple neighborhood matrix specifications for spatio-temporal model fitting: A case study on covid-19 data. *Stochastic Environmental Research and Risk Assessment*, 36(1), 271-282.
- Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the Total Environment*, 728, 1-7.

- Cheong, Y. L., Ghazali, S. M., Che Ibrahim, M. K. B., Kee, C. C., Md Iderus, N. H., Gill, B. S., Lee, F. C. H., & Lim, K. H. (2022). Assessing the spatiotemporal spread pattern of the COVID-19 Pandemic in Malaysia. *Frontiers in Public Health*, 10, 301.
- Coşkun, H., Yıldırım, N., & Gündüz, S. (2021). The spread of COVID-19 virus through population density and wind in Turkey cities. *Science of the Total Environment*, 751, 141663.
- Dayun, K., Huynh, C., Jong, H., & Jungsoon, C. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases*, 94, 96-102.
- Dong, E., Ratcliff, J., Goyea, T. D., Katz, A., Lau, R., Ng, T. K., Garcia, B., Bolt, E., Prata, S., & Zhang, D. (2022). The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: Data collection process, challenges faced, and lessons learned. *The Lancet Infectious Diseases*, 22(12), 370-376.
- DOSM, (2021). Department of Statistics Malaysia (DOSM): Key findings population and housing census of Malaysia, 2020 version 2.0. Retrieved from <https://open.dosm.gov.my/>
- Duckett, S. J., & Sutton, B. (2021). On entering Australia's third year with COVID-19. *Medical Journal of Australia*, 215, 509-510.
- Dutta, I., Basu, T., & Das, A. (2021). Spatial analysis of COVID-19 incidence and its determinants using spatial modeling: A study on India. *Environmental Challenges* 4, 1-10.
- El Deeb, O. (2021). Spatial autocorrelation and the dynamics of the mean center of COVID-19 infections in Lebanon. *Frontiers in Applied Mathematics Statistics*, 6, 620064.

- Fallah-Aliabadi, S., Fatemi, F., Heydari, A., Khajehaminian, M. R., Lotfi, M. H., Mirzaei, M., & Sarsangi, A. (2022). Social vulnerability indicators in pandemics focusing on Covid-19: A systematic literature review. *Public Health Nursing, 39*(5), 1142-1155.
- Faziera, Y. N., Elizabeth, E. A., Danggat, C., & Tarmiji, M. (2020). Coronavirus (COVID-19): Density risk mapping using Population and Housing Census of Malaysia 2010: Coronavirus (COVID-19): Pemetaan risiko kepadatan menggunakan data Banci Penduduk dan Perumahan Malaysia 2010. *Geografi, 8*(2), 21-47.
- Franch-Pardo, I., Desjardins, M. R., Barea-Navarro, I., & Cerdà, A. (2020). Spatial analysis and GIS in the study of COVID-19. A review. *Science of the Total Environment, 739*, 140033.
- Franch-Pardo, I., Desjardins, M. R., Barea-Navarro, I., & Cerdà, A. (2021). A review of GIS methodologies to analyze the dynamics of COVID-19 in the second half of 2020. *Transactions in GIS, 25*(5), 2191-2239.
- 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 Public Health, 19*(4), 1-13.
- Gargoum, S. A., & Gargoum, A. S. (2021). Limiting mobility during COVID-19, when and to what level? An international comparative study using change point analysis. *Journal of Transport & Health, 20*, 101019.
- Goscé, L., Phillips, A., Spinola, P., Gupta, R. K., & Abubakar, I. (2020). Modelling SARS-COV2 spread in London: approaches to lift the lockdown. *Journal of Infection, 81*, 260-265.

- Grether, J. M., & Mathys, N. A. (2010). Is the world's economic centre of gravity already in Asia? *Area*, 42(1), 47-50.
- Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics*, 38, 1-10.
- Gupta, D., Biswas, D., & Kabiraj, P. (2022). COVID-19 outbreak and Urban dynamics: Regional variations in India. *Geo Journal*, 87(4), 2719-2737.
- Huang, Z. (2021). Spatiotemporal evolution patterns of the COVID-19 pandemic using space-time aggregation and spatial statistics: A global perspective. *International Journal of Geo-Information*, 10(8), 1-17.
- Iderus, N. H. M. , Singh, S. S. L., Ghazali, S. M., Ling, C. L., Vei, T. C., Md Zamri, A. S. S., Jaafar, N. A., Ruslan, Q., Jaghfir, N. H. A., & Gill, B. S. (2022). Correlation between population density and COVID-19 cases during the third wave in Malaysia: Effect of the delta variant. *International Journal of Environmental Research Public Health*, 19(12), 1-17.
- Jayaraj, V. J., Rampal, S., Ng, C. W., & Chong, D. W. Q. (2021). The epidemiology of COVID-19 in Malaysia. *The Lancet Regional Health-Western Pacific*, 17, 100295.
- Kan, Z., Kwan, M. P., Huang, J., Wong, M. S., & Liu, D. (2021). Comparing the space-time patterns of high-risk areas in different waves of COVID-19 in Hong Kong. *Transactions in GIS*, 25(6), 2982-3001.
- Kang, D., Choi, H., Kim, J. H., & Choi, J. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases*, 94, 96-102.
- Killick, R., & Eckley, I. (2014). Changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58(3), 1-19.

- Kumpel, E., Billava, N., Nayak, N., & Ercumen, A. (2022). Water use behaviors and water access in intermittent and continuous water supply areas during the COVID-19 pandemic. *Journal of Water Health*, 20(1), 139-148.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 172(9), 577-582.
- Li, H., Li, H., Ding, Z., Hu, Z., Chen, F., Wang, K., Peng, Z., & Shen, H. (2020). Spatial statistical analysis of coronavirus disease 2019 (Covid-19) in China. *Geospatial Health*, 15(1), 11-17.
- Li, X., Zhou, L., Jia, T., Peng, R., Fu, X., & Zou, Y. (2020). Associating COVID-19 severity with urban factors: A case study of Wuhan. *International Journal of Environmental Research Public Health Reports*, 17(18), 1-20.
- Liu, Y., He, Z., & Zhou, X. (2020). Space-time variation and spatial differentiation of COVID-19 confirmed cases in Hubei Province based on extended GWR. *International Journal of Geo-Information*, 9(9), 1-17.
- Lupton, D. (2021). Contextualising COVID-19: Sociocultural perspectives on contagion. In *the COVID-19 Crisis*: Routledge (14-24).
- Ma, Q., Gao, J., Zhang, W., Wang, L., Li, M., Shi, J., Zhai, Y., Sun, D., Wang, L., & Chen, B. (2021). Spatio-temporal distribution characteristics of COVID-19 in China: a city-level modeling study. *BMC Infectious Diseases*, 21, 1-14.
- Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A., & Pilla, F. (2021). Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. *Sustainable Cities Society*, 68, 1-15.

- Mitze, T., & Kosfeld, R. (2022). The propagation effect of commuting to work in the spatial transmission of COVID-19. *Journal of Geographical Systems*, 24(1), 5-31.
- Muus, C., Luecken, M. D., Eraslan, G., Sikkema, L., Waghray, A., Heimberg, G., Kobayashi, Y., Vaishnav, E. D., Subramanian, A., & Smillie, C. (2021). Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nature Medicine*, 27(3), 546-559.
- Osayomi, T., Adeleke, R., Akpoterai, L. E., Fatayo, O. C., Ayanda, J. T., Moyin-Jesu, J., Isioye, A., & Popoola, A. (2021). A geographical analysis of the African COVID-19 paradox: putting the poverty-as-a-vaccine hypothesis to the test. *Earth Systems Environmental Challenges*, 5(3), 799-810.
- Phang, P., Taib, N. A., Safii, R., & Labadin, J. (2021). Piecewise linear modelling and change-point analysis of COVID-19 outbreak in Malaysia. *Journal of Physics: Conference Series*, 1988 (2021), 1-13.
- Rahman, M. R., Islam, A. H., & Islam, M. N. (2021). Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. *Modeling Earth Systems Environment*, 7, 2059-2087.
- Ramírez-Aldana, R., Gomez-Verjan, J. C., & Bello-Chavolla, O. Y. (2020). Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. *PLOS Neglected Tropical Diseases*, 14(11), e0008875.
- Rampal, L., & Liew, B. (2021). Malaysia's third COVID-19 wave-a paradigm shift required. *The Medical Journal of Malaysia*, 76(1), 1-4.

- Rampal, L., Liew, B.S., Choolani, M., Ganasegeran, K., Pramanick, A., Vallibhakara, S.A., Tejavivaddhana, P., & Hoe, V.C. (2020). Battling COVID-19 pandemic waves in six South-East Asian countries: A real-time consensus review. *Medical Journal of Malaysia*, 75(6), 613-625.
- Rampal, L., & Seng, L. B. (2020). Coronavirus disease (COVID-19) pandemic. *Medical Journal of Malaysia*, 75(2), 95-97.
- Reyes, R., Ahn, R., Thurber, K., & Burke, T. F. (2013). Urbanization and infectious diseases: general principles, historical perspectives, and contemporary challenges. *Challenges in infectious diseases*, 123-146.
- Ribeiro, H. V., Sunahara, A. S., Sutton, J., Perc, M., & Hanley, Q. S. (2020). City size and the spreading of COVID-19 in Brazil. *PLOS One*, 15(9), e0239699.
- Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., Goodman, S., Grimsley, G., Layko, R., & Melville, G. (2020). GeoBoundaries: A global database of political administrative boundaries. *PLOS One*, 15(4), e0231866.
- Sarkar, S. K., Ekram, K. M. M., & Das, P. C. (2021). Spatial modeling of COVID-19 transmission in Bangladesh. *Spatial Information Research*, 29, 1-12.
- Shah, A. U. M., Safri, S. N. A., Thevadas, R., Noordin, N. K., Abd Rahman, A., Sekawi, Z., Ideris, A., & Sultan, M. T. H. (2020). COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International Journal of Infectious Diseases*, 97, 108-116.
- Sun, F., Matthews, S. A., Yang, T. C., & Hu, M. H. (2020). A spatial analysis of the COVID-19 period prevalence in US counties through June 28, 2020: where geography matters? *Annals of Epidemiology*, 52, 54-59.

- Tammes, P. (2020). Social distancing, population density, and spread of COVID-19 in England: a longitudinal study. *British Journal of General Practice (BJGP) Open*, 4(3), 1-5.
- Troppy, S., Wilt, Grete E., Whiteman, A., Hallisey, E., Crockett, M., Sharpe, J. D., Haney, G., Cranston, K., & Klevens, R. M. (2021). Geographic associations between social factors and SARS-CoV-2 testing early in the COVID-19 Pandemic, February-June 2020, Massachusetts. *Public Health Reports*, 136(6), 765-773.
- Tu, J., & Xia, Z. G. (2008). Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of the Total Environment*, 407(1), 358-378.
- Ulimwengu, J., & Kibonge, A. (2021). Spatial spillover and COVID-19 spread in the US. *BMC Public Health*, 21(1), 1-11.
- Wang, Y. Y., Liang, Z., Ding, J. Q., Sun, F. Y., & Li, S. C. (2022). Identify the natural and socio-economic influencing factors of the new coronavirus pneumonia (COVID-19) incidence rates in Chinese cities. *Zhongguo Huanjing Kexue China Environmental Science*, 1418-1418.
- Wong, D. W., & Li, Y. (2020). Spreading of COVID-19: Density matters. *PLOS One*, 15(12), 1-16.
- Wong, H. S., Hasan, M. Z., Sharif, O., & Rahman, A. (2023). Effect of total population, population density and weighted population density on the spread of Covid-19 in Malaysia. *PLOS One*, 18(4), e0284157.
- World Health Organization. (2021). World Health Organization Coronavirus (COVID-19) Dashboard. Retrieved from <https://covid19.who.int/>

- Wu, W., & Zhang, L. (2013). Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in north-eastern Puerto Rico. *Applied Geography*, 37, 52-62.
- Wu, X., & Zhang, J. (2021). Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR). *Environmental Science and Pollution Research*, 28, 43732-43746.
- Xie, Z., Qin, Y., Li, Y., Shen, W., Zheng, Z., & Liu, S. (2020). Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. *Science of the Total Environment*, 744, 140929.
- Yang, L., Han, Y., Nilsson-Payant, B. E., Gupta, V., Wang, P., Duan, X., Tang, X., Zhu, J., Zhao, Z., & Jaffré, F. (2020). A human pluripotent stem cell-based platform to study SARS-CoV-2 tropism and model virus infection in human cells and organoids. *Cell Stem Cell*, 27(1), 125-136.
- Yi, H., Ng, S. T., Farwin, A., Pei Ting Low, A., Chang, C. M., & Lim, J. (2021). Health equity considerations in COVID-19: Geospatial network analysis of the COVID-19 outbreak in the migrant population in Singapore. *Journal of Travel Medicine*, 28(2), 1-8.
- Zakaria, S., Zaini, N. E., Malik, S. M. A., & Alwi, W. S. W. (2021). Exploratory spatial data analysis (ESDA) on COVID-19 cases in Malaysia. *Jurnal Teknologi*, 83(6), 83-94.
- Zamri, A. S. S. M., Singh, S., Ghazali, S. M., Heng, L. C., Dass, S. C., Aris, T., Ibrahim, H. M., & Gill, B. S. (2021). Effectiveness of the movement control measures during the third wave of COVID-19 in Malaysia. *Epidemiology Health*, 43, 1-8.

- Zhang, M., Wang, S., Hu, T., Fu, X., Wang, X., Hu, Y., Halloran, B., Li, Z., Cui, Y., & Liu, H. (2022). Human mobility and COVID-19 transmission: A systematic review and future directions. *Annals of GIS*, 28(4), 501-514.
- Zhou, C., Tao Pei, An Zhang, Yunyan Du, Bin Luo, Zhidong Cao, Juanle Wang, Wen Yuan, Yunqiang Zhu, Ci Song, Jie Chen, Jun Xu, Fujia Li, Ting Ma, Lili Jiang, Fengqin Yan, Jiawei Yi, Yunfeng Hu, Yilan Liao,, & Xiao., H. (2020). COVID-19: Challenges to GIS with big data. *Geography and Sustainability*, 1(1), 77-87.
- Zhu, S., Tao, J., Gao, H., & He, D. (2020). Age, source, and future risk of COVID-19 infections in two settings of Hong Kong and Singapore. *BMC Research Notes*, 13, 1-3.