Modelling and analysis of COVID-19 outspread at micro-levels using spatial autocorrelation: Case of eThekwini.

By

Samukelisiwe Ngubane (1870201)

Supervisors: Dr Dorman Chimhamhiwa and Prof. Elhadi Adam

A research report submitted in partial fulfilment of the requirements for the degree in Master of Science in GIS and Remote Sensing

Faculty of Science

University of the Witwatersrand

Johannesburg, South Africa

ABSTRACT

The alarming effects of the COVID-19 pandemic on different socio-economic spheres have been felt across the globe. These destructive effects have prompted plenty of research to understand and control the coronavirus pandemic. Notably, one strategic method of mitigating the effects of the coronavirus epidemic has been the utilisation of spatial and geostatistical models to gain insights into the potential predictors of the prevalence of the coronavirus.

Considering the above, it was the aim of this study to explore the use of advanced geospatial modelling and analysis techniques, including Moran's *I*, spatial error models, spatial lag models, MGWR, and GWR for analysing and modelling the settlement level determining factors of COVID-19 incidence within the eThekwini Metro to inform effectual micro-level planning. Notably, the lack of micro-level modelling of COVID-19 prevalence predictors also motivated the undertaking of this study.

To the above aim, the objectives of the research were to utilise spatial autocorrelation to map the granular level COVID-19 spatial distribution over the 3rd wave in the eThekwini Metro, compare the applicability of global and local models in analysing and modelling micro-level COVID-19 incidence, analyse the spatial dependence of the occurrence of COVID-19 on local level variables through Moran's *I* and to spatially model the effects of significant local-level determinants on COVID-19.

The incidence of COVID-19 cases for the 3rd wave, which was from the 2nd of May 2021 to the 11th of September 2021, was analysed and modelled. The Moran's *I* result illustrated that COVID-19 incidence within the eThekwini settlement places had a positive spatial autocorrelation, with a Moran's *I* value of 0.14 and a p-value of 0.00. Also, the MGWR model's local R^2 value was greater (72.5%) as compared to the other models. Moreover, economic wellness score, the sum of TB cases and population density came out as the significant determining factors of settlement level incidence of COVID-19. This research report offers a great foundation for gaining insights into the applicability of advanced geospatial models in guiding targeted COVID-19 interventions at lower levels.

ACKNOWLEDGEMENTS

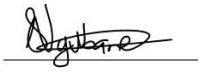
I want to convey my earnest appreciation to my supervisors Dr. Dorman Chimhamhiwa and Prof. Elhadi Adam for giving meaningful insights, data, supervision, and support during the course of my research. My deepest appreciation to the Right to Care GIS department for their sharing of knowledge, skills, and expertise during my study. Their warm help helped shape my research.

Finally, I am eternally thankful to my dear family for their unwavering support, reassurance, love, and prayers during the course of my studies.

Plagiarism Declaration Form

I, Samukelisiwe Ngubane (Student number: 1870201), hereby declare that:

- The "Modelling and analysis of COVID-19 outspread at micro-levels using spatial autocorrelation: Case of eThekwini" is my own work.
- All the sources quoted in this research have been re-rewritten andacknowledged in the reference list.
- This research report has never been submitted at any other university forexamination or any degree.
- This research report does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the sourcebeing detailed in the research report and in the Reference List section.

Signature:  Date: 17-09-2024

Samukelisiwe Ngubane

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

COVID-19 – Coronavirus Disease 2019

DHIS – District Health Information Software

GAM – Geographic Analysis Machine

GIS – Geographical Information Systems

OLS – Ordinary Least Squares

GWR – Geographically Weighted Regression

GLM – General Linear Model

MGWR – Multiscale Geographically Weighted Regression

MM – Metropolitan Municipality

VIF – Variance Inflation Factors

NU – Non-Urban

NDoH – National Department of Health

SARSCOV-2 – Severe Acute Respiratory Syndrome Coronavirus 2

SEM – Spatial Error Model

SLM – Spatial Lag Model

EVDS – Electronic Vaccination Data System

CHAPTER 1

General Introduction

1.1. Introduction

Epidemiology, which concentrates on the analysis of the distribution of communicable diseases along with their patterns and movement, has a long research history, particularly for researchers within the fields of geography and medicine, with the earliest research going back to the fifth century B.C. when epidemiological analysis was employed in investigating and mitigating the widespread occurrence or outbreak of mumps in an island located in the north-eastern region of Greece known as Thasos (Barrett, 2000; Krieger, 2003; Tsoucalas *et al*., 2013; Editorial, 2018; Cromley, 2019). From these earlier times, geography has grown significantly and has become a crucial component in the expansion of epidemiology and a key contributor in the establishment and development of what is now known as spatial epidemiology or geographical medicine, which depends on place or location to make sense of aetiology, conduct epidemiological analysis and visualise the spread of diseases across various spatiotemporal extents (Barrett, 2000; Krieger, 2003; Fatima *et al*., 2021; Shrestha *et al*., 2022).

The above is well-demonstrated when looking at the case of the mumps epidemic in Thasos, where the distribution, hotspots and diffusion patterns of the mumps epidemic were mapped through the use of geography, which enabled the discovery of harbours as the main entry points of the outbreak (Tsoucalas *et al*., 2013). In a similar way, fundamental geospatial thinking was utilised by Snow (1856) to physically map the locations of the deaths and incidence of the cholera epidemic which occurred in London in 1854 and led to the discovery of the contaminated Broad Street water pumps as the root causes of the cholera outbreak (Snow, 1856; Fatima *et al*., 2021). In recent times, geospatial technologies, such as geographic information systems (GIS), have seen several ground-breaking developments, which have greatly improved geographical medicine by enhancing the spatiotemporal quantification, modelling, and presentation of epidemics and the risk factors linked to the epidemics (Cromley, 2019; Fatima *et al*., 2021; Shrestha *et al*., 2022). Due to the advancements

in GIS technologies, monitoring and control of disease endemics have been enhanced (Cromley, 2019).

Markedly, the geospatial analysis of the most recent epidemic, medically known as SARS-CoV-2, which is an acronym for the severe acute respiratory syndrome coronavirus 2, generally referred to as COVID-19 or coronavirus, is evidence of the epidemiological analysis potential of GIS (Hassaan *et al*., 2021; Ahasan *et al*., 2022). Notably, the first incidence of the coronavirus was towards the end of 2019 in the city of Wuhan, China and was later, in March 2020, acknowledged as a large-scale pandemic by the World Health Organisation following its increased severity resulting from the lack of a preventative vaccine and its rapid transmission across the world mostly through contact with contaminated air, droplets and in some instances, contaminated surfaces (Castro *et al*., 2021; Manda *et al*., 2021; Hassaan *et al*., 2021). There have been over 6,9 million deaths around the globe due to the COVID-19 pandemic and several economies, healthcare sectors and social industries across the world have been severely strained due to the pandemic (Ahasan *et al*., 2022). The above is especially true for South Africa, which, at the time of the study, has had fatalities surpassing 100,000 since the confirmation of the first COVID-19 incidence in KwaZulu-Natal on March 5 2020, and has experienced an alarming deterioration of various economies and increased strain on healthcare systems as a result of the epidemic (Broadbent *et al*., 2020; Manda *et al*., 2021).

In addition to the above, the presence of most of the determining factors of COVID-19 incidence and morbidity in South Africa, including overcrowding, pollution, high humidity, comorbidities, the highest proportion of old people in Africa, 9.1%, strained healthcare systems and poverty further aggravate the detrimental effects of the coronavirus epidemic in South Africa (Statistics South Africa, 2020; Manda *et al*., 2021). Bearing in mind the devastating impact of the virus, it is important to gain insights into the diffusion patterns and spatial spread of the pandemic to inform pandemic control strategies (Hassaan *et al*., 2021). In this respect, geospatial analysis and modelling have been crucial, as they provide GIS systems and tools which allow for the mapping, modelling and visualisation of the coronavirus spatial distribution, hotspots, determinants, and vaccination sites along with prediction of COVID-19 incidence, which enhance the strategic response of the National Department of Health (NDOH) to the virus (Gibson and Rush, 2020; Franch-Pardo *et al*., 2020).

Advanced geospatial models, including Markov regression and spatial autocorrelation, have been applied to map the descriptive variables of COVID-19 along with its spatial distribution (Saffary, *et al*., 2020; Shariati *et al*., 2020). Notably, while most advanced geospatial models enhance understanding of COVID-19 spread, they have mostly been utilised at relatively large spatiotemporal scales leading to the lack of accurate granular level modelling (Pillay *et al*., 2021). Subsequently, there exists a need for the local level utilisation of advanced spatial models to understand COVID-19 incidence (Fatima *et al*., 2021). The above is specifically true for South Africa where rudimentary geospatial mapping and modelling techniques, including cluster or hotspot analysis and choropleth maps, have been predominantly employed in studies aiming to acquire an understanding of the outspread of COVID-19 (Gibson and Rush, 2020; Mukandavire *et al*., 2020). In the few cases where advanced geospatial modelling has been utilised, it has been undertaken at relatively large spatial scales, such as at the regional scale, thus not providing a clear view of the micro-level dynamics of COVID-19 spread (Mukandavire *et al*., 2020). Subsequently, it is essential that the spatiotemporal modelling of COVID-19 be improved by employing enhanced spatial models, such as spatial autocorrelation, to map and model the coronavirus at finer spatial scales, including wards and settlements.

1.2. Problem statement

As mentioned briefly in the above sections, GIS-based mapping and analysis have been utilised extensively for granular-level monitoring, planning, targeting and implementation of various response strategies at different spatial levels since the beginning of COVID-19 (Gibson and Rush, 2020; Ahasan *et al*., 2022). The above-mentioned response strategies include planning for field activities, choosing areas where COVID-19 testing and vaccination teams should be deployed, responding to hotspot outbreaks and contact tracing (Gibson and Rush, 2020; Mukandavire *et al*., 2020; Shariati *et al*., 2020). Nevertheless, at the time of completion of this research, a majority of South African studies aiming to model spatiotemporal dynamics of COVID-19 employed non-robust and basic spatial approaches, including dot density mapping and choropleth maps, which do not offer micro-level information regarding the spatial distribution of COVID-19 (Shariati *et al*., 2020; Fatima *et al*., 2021). Resultantly, existing spatial models and analysis of COVID-19 spread do not guide granular level decision-making, including decisions for settlements and wards, as these models are

often too abstract or isolated to fully rationalise the distribution patterns of COVID-19 at micro-levels.

The above is particularly true for the Metropolitan Municipality (MM) of eThekwini, which has been a COVID-19 hotspot or epicentre and has only been targeted for abstract and/or non-robust spatial modelling, including mapping of COVID-19 clusters and confirmed cases, and analysis of COVID-19 spread at the district scale, which has not provided greater understanding into the dynamics of the incidence of COVID-19 and its spread at the finest geographic scales (Maharaj and Reddy, 2020; Mbambo and Agbola, 2020; Pillay *et al*., 2021). Accordingly, the decision-making of the National Department of Health and response strategies at the finest spatial levels have been misled due to existing spatial models not fully explaining the geographic distribution of COVID-19 along with the explanatory variables at a micro-level. Consequently, this study explores the utilisation of advanced spatial analysis and modelling, including GWR and spatial autocorrelation to map and model the spread of COVID-19. The above will enhance spatial understanding and visualisation of COVID-19 distribution at granular levels, decrease bias of visual perceptions, inform effectual COVID-19 response strategies and eventually contribute to improved GIS-based outcomes in geographical medicine (Fatima *et al*., 2021). The above is motivated by the absence of robust granular level information regarding the geographic distribution of COVID-19 spread at settlement, and ward levels in South Africa, specifically in eThekwini.

1.3. Aims and objectives.

With the above in mind, this study aims to investigate the pertinency of advanced spatial models, including MGWR and spatial autocorrelation in analysing and modelling COVID-19 spread at much finer spatial levels.

The particular objectives of the study are:

1. To map the micro-level spatial distribution of COVID-19 over the 3rd wave in eThekwini using spatial autocorrelation (Moran's *I* statistic).
2. To compare the applicability of global and local models in analysing and modelling micro-level COVID-19 incidence.
3. To analyse the spatial dependence of COVID-19 incidence on local level variables using spatial autocorrelation.

4. To spatially model the effects of significant local-level determinants on COVID-19.

1.4. Research report structure

The research report comprises six chapters. The first chapter provides a brief background and rationale for undertaking the research and provides an outline of the research objectives and main aim. The second chapter is concerned with providing a review of the literature and outlines crucial theoretical concepts that are related to this research, including the history of spatial epidemiology, techniques for disease mapping, and applications of spatial modelling in COVID-19. The third chapter characterizes the research study area and describes in depth the research materials along with the methodology employed in the study to reach the objectives stated in the first chapter. The research results are then presented and described in the fourth chapter while the fifth chapter offers a thorough analysis and discussion of the results and research study limitations. Lastly, the conclusions drawn from the research and recommendations are detailed in chapter six.

CHAPTER 2

Literature Review

2.1. Spatial epidemiology and disease mapping

Epidemiology as it is known today, is a product of various developments and advancements in epidemiology and related fields (Krieger, 2003). Notably, the increased use of GIS and spatial statistics to understand pandemics is among the advancements that have shaped modern epidemiology (Cromley, 2019; Shrestha *et al*., 2022). Additionally, the understanding amongst epidemiological scholars is that viewing epidemics and endemics as rigid articles with fixed symptoms that are robust to changes and influences from peripheral factors is outdated and fails to take into account geography or location, which is an important subject in understanding aetiology (Cromley, 2019; Shrestha *et al*., 2022). Subsequently, the utilisation of GIS tools and systems in epidemiology has increased considerably within the past six decades or so, due to the developments and advancements in GIS-based tools which offer real-time monitoring, projection, and straightforward and clear visualisation of infectious disease trends (Cromley, 2019; Ahasan *et al*., 2022; Shrestha *et al*., 2022).

Notably, in the early days of epidemiology, statistical analysis and scientific observation (laboratory analysis) were at the core of understanding aetiology and disease patterns (Susser, 1989; Morabia, 2015). The above is evidenced in various studies conducted in the late eighteenth and early nineteenth centuries utilising statistical analysis and scientific observation to understand disease distribution, risk factors and causes (Susser, 1989; Morabia, 2015; Editorial, 2018). Amongst such studies is that of John Graunt who employed quantitative analysis and basic statistics to estimate infant mortality and quantify patterns of births and diseases along with the impact of an epidemic fever in 1662 on population size (Kargon, 1963; Connor, 2022). John Graunt's work resulted in the development and widespread utilisation of vital statistics in epidemiology, which greatly improved the quantitative understanding of epidemics and endemics (Kargon, 1963). Nevertheless, statistics and scientific observation could not still fully analyse and effectively visualise the spatial and temporal nature of disease occurrence (Clarke *et al*., 1996). Consequently, spatial epidemiology was developed to enrich epidemiological and public health outputs through assessing spatial patterns of diseases and exploring associations between

disease occurrence, and temporal and spatial variations (Stevenson, 1965; Morabia, 2015; Shrestha *et al*., 2022).

Noticeably, the earliest applications of spatial analysis in epidemiology were centred around producing disease spot maps to identify risk factors and describe spatial variations in the occurrence of pandemics to inform aetiological hypotheses and preventative action (Stevenson, 1965; Rytkönen, 2004). For instance, Dr Robert Baker produced a cholera map showing the spatial distribution of cholera cases in Leeds, England in 1833 and noted that areas with poor sanitation corresponded with a higher incidence of cholera while Valentina Seaman mapped and illustrated the spatial correlation between yellow fever deaths and the locations of waste dumps around New York in 1798 (Stevenson, 1965; Krieger, 2003; Lin and Wen, 2022).

2.2. Advancements in spatial modelling techniques

Over time, technological advancements, including advances in GIS technologies, have improved disease mapping for epidemiology and introduced advanced geospatial analysis techniques for understanding infectious diseases (Clarke *et al*., 1996; Krieger, 2003). The development of geodatabases and cloud-based GIS storage facilities, which enable large health datasets to be stored centrally in servers where they can be accessed securely and readily for epidemiological analysis, is also amongst some of the important GIS tools and systems advancements.

In addition to the above, recently developed software systems and packages, including ArcLogistics and BodyViewer, offer enhanced and innovative views and insights about clinical data (Mishra and Kumar, 2021). As a result of the above, the analysis of disease hotspots and risk factors is improved (Mishra and Kumar, 2021). To illustrate, in earlier studies, the highly advanced GIS tool known as the Geographic Analysis Machine (GAM) developed in 1987 by Openshaw was utilised in mapping significant clusters of leukaemia and successfully showed that the susceptibility of leukaemia and other cancers increased substantially with increasing proximity to nuclear facilities (Mishra and Kumar, 2021). Noticeably, when compared to traditional methods of mapping leukaemia, GAM also proved to be an economical alternative; as is the case with several other GIS solutions (Krieger, 2003; Shrestha *et al*., 2022).

Moreover, advancements in geospatial software systems have also led to the improvement of GIS-based tools and geospatial analysis approaches (Cromley, 2019; Ahasan *et al*., 2022). To put the above into context, Mahara *et al*. (2018) combined statistics with GIS through advanced spatial autocorrelation regression models and mapped the spatial spread of Tuberculosis along with its predictor variables. The above study contributed to the understanding of TB prevalence and its association with socio-economic factors in Beijing, China (Mahara *et al*., 2018).

2.3. Spatial analysis and modelling of COVID-19

In recent times, the use of geospatial and geostatistical tools for understanding and responding to the coronavirus pandemic has increased significantly due to the applicability and previous success of these tools and methods in other epidemiological studies as stated above (Mahara *et al*., 2018). Geospatial modelling and analysis have been utilised in several studies concerning COVID-19, including those aiming to understand the explanatory variables of COVID-19 distribution (Kindi *et al*., 2021). For example, Kindi *et al*. (2021) explored the capabilities of a general linear model (GLM) and GWR for analysing socioeconomic determining variables of COVID-19 incidence in Oman and noted that the spatial model (GWR) provided a more robust model of the relationship between COVID-19 incidence and predictor variables. In addition to the above, the study also found that age groups, population density, and number of hospital beds were some of the significant predictors of the occurrence of COVID-19 (Kindi *et al*., 2021). The above results were obtained through the utilisation of spatial and statistical models (Kindi *et al*., 2021). Also, the results from the above study were accurate and demonstrated the applicability of advanced geospatial models in analysing the COVID-19 pandemic, notwithstanding the limitation of the relatively abstract or macro spatial scale used in the study (Kindi *et al*., 2021).

In a study comparable to the above, geospatial analysis in the form of GWR, spatial autocorrelation and OLS models was utilised by Ahasan *et al*. (2022) to map and model the explanatory variables of COVID-19 prevalence and mortality in African countries. This research study highlighted overcrowding, health status, pollution, and vaccination access as some of the key determinants of the spatial distribution of COVID-19 across African countries, including, Egypt, Libya, and Namibia (Ahasan *et al*., 2022). Nevertheless, the above study was not without limitations as the spatial modelling and mapping were undertaken at a relatively large spatial scale and

subsequently, could not guide decision-making at granular spatial levels (Shariati *et al*., 2020; Ahasan *et al*., 2022).

In COVID-19 incidence studies, Saffary *et al*. (2020) used Queen's contiguity spatial weights and Moran's *I* to acquire an understanding of the spatial relationship between COVID-19 incidence and mortality and populations at county levels. The study indicated that COVID-19 incidence and mortality were substantially increased by limited health facilities and ICU beds, high population density, and the presence of comorbidities (Saffary *et al*., 2020). In addition to the above, through the use of spatial models, the study found that Hispanic and Black communities generally lacked access to quality healthcare systems (Saffary *et al*., 2020). Moreover, the above study was able to provide an enhanced understanding of the association between the characteristics of healthcare systems and epidemics (Saffary *et al*., 2020).

In addition to the above, it is notable that descriptive statistics and correlation analysis have been widely utilised for COVID-19 spatial modelling and have produced valuable results, as shown in the study conducted by Shariati *et al*. (2020), which utilised Moran's *I* along with cluster analysis to map global COVID-19 hotspots and to offer valuable information about the availability of resources for responding to the COVID-19 pandemic (Shariati *et al*., 2020). Notably, Shariati *et al*. (2020) used hotspot analysis and Anselin Local Moran's *I* to accurately identify and visualise low and high clusters of coronavirus in all the countries impacted by the pandemic as of April 2020. Additionally, the study provided valuable insights into the countries that required additional resources and monitoring due to a high number of cumulative COVID-19 cases and deaths, which were identified through spatial modelling (Shariati *et al*., 2020). Additionally, Sangkham *et al*. (2021) used descriptive spatial statistics and correlation analysis in the form of Spearman and Kendall's coefficients to examine the geographic association between COVID-19 incidence and meteorological factors, such as air pollution in Bangkok. This study highlighted that COVID-19 incidence was aggravated by high levels of air pollution and produced useful information for decision-making and pandemic control in Thailand (Sangkham *et al*., 2021).

In most instances, local spatial regression models have been the preferred models for spatial analysis of COVID-19 due to their ability to model spatial autocorrelation using spatially lagged error terms, making it easier to observe the local-level spatiotemporal

dynamics of COVID-19 and other phenomena (Mollalo *et al*., 2020). This is especially true for models such as GWR and MGWR, which model local level variations in data (Mollalo *et al*., 2020; Dutta *et al*., 2021). Notably, GWR and MGWR often have the advantage of providing an accurate representation of relationships between variables at the lowest level (Akinwumiju *et al.,* 2022; He *et al.,* 2023). Consequently, these spatial regression models are best suited for use when there is a need to consider spatial dependencies when analysing relationships between variables, as they account for local variation and do not assume a single global relationship between variables (Mollalo *et al*., 2020; Dutta *et al*., 2021; He *et al.,* 2023).

For instance, GWR captures local variations in the relationships between explanatory and dependent variables, making it suitable for cases where there is evidence of spatially varying relationships between dependent and explanatory variables, such as the spread of COVID-19 and socioeconomic variables, such as population density accessibility of healthcare facilities (Saffary *et al*., 2020; Akinwumiju *et al.,* 2022). Similarly, MGWR accounts for spatial heterogeneity and allows for different variables to be analysed at varying spatial scales as it assumes that the influence of explanatory variables on the dependent variable changes across different spatial scales (Dutta *et al*., 2021). However, these local spatial regression models are computationally intensive and require careful selection of the neighbourhood distance or bandwidth for the explanatory variables to perform optimally (Brunsdon, 1996; Chandra and Sharma, 2024). As a result, they are often not ideal for large datasets and for modelling sparse data.

Comparatively, spatial Bayesian hierarchal models, which use Bayesian statistical methods with multiple layers representing different sources of variation, are more suitable for the spatial analysis of large and complex datasets (Wang *et al.,* 2024). This is due to their ability to handle various hierarchal relationships between variables and quantify uncertainty, especially when spatial heterogeneity is present (Ramírez-Aldana *et al.,* 2020; Wang *et al.,* 2024). Therefore, Bayesian hierarchical models are suitable for accurately identifying COVID-19 hotspots based on multiple explanatory variables over large spatial scales and predicting COVID-19 hotspots with limited data (Castro *et al*., 2021; Wang *et al.,* 2024).

It is noticeable that spatial analysis and modelling of various aspects of COVID-19 have also been performed in South Africa. However, most of the published studies did not utilise advanced geospatial methods and mainly studied the spatial distribution of COVID-19 at macro-levels. However, there are a few studies which investigated COVID-19 through integrating foundational spatial analysis techniques with advanced geospatial approaches. To illustrate, Mukandavire *et al*. (2020) fitted the Bayesian Markov Chain and a *SEIR* model (which are both advanced geospatial models) with the cumulative number of confirmed COVID-19 cases (from the beginning of the epidemic to March 2020) in South Africa covering to quantify and predict the magnitude of the pandemic while estimating the desirable efficacy of vaccines to curtail the spread of the pandemic. The above study was able to show that to adequately respond to the pandemic in South Africa, an efficacy of 70% for vaccines would be required if there was high vaccination uptake (Mukandavire *et al*., 2020). Additionally, the study projected a drastic surge in the sum of confirmed COVID-19 cases throughout the country during the winter of 2020, which was the case. In the same year, Gibson and Rush (2020) studied the practicality of social distancing within the Cape Town informal settlements using descriptive statistics, distance and cluster analysis and buffer analysis. Through the use of spatial analysis, the study was able to note the impracticality and lack of efficiency of social distancing within informal settlements, especially in the absence of national lockdown regulations (Gibson and Rush, 2020).

In addition to the above, spatial prediction and analysis of COVID-19 spread post-lockdown was also undertaken by Arashi *et al*. (2020). Through mapping COVID-19 hotspots at a provincial level, computing spatial autocorrelation of COVID-19, and using logistic regression to predict COVID-19 cases for each province, the above study illustrated the positive impact of lockdown regulations between March 21 to April 25 2020 and showed that COVID-19 spread depended more on adjacency rather than distance to the epicentre (Arashi *et al*., 2020). Although the outcomes of the study were limited by the scale of the research, the outcomes were still important for informing policy and decision-making (Rytkönen, 2004).

Noticeably, spatial modelling of COVID-19 has also been conducted for eThekwini MM. (Pillay *et al*., 2021). To illustrate, Pillay *et al*. (2021) analysed the presence of the coronavirus viral RNA in wastewater by utilising descriptive statistics and subsequently

measured the number of COVID-19 cases from the levels of SARS-CoV-2 in wastewater. Noticeably, the study was conducted for 14 weeks in four wastewater plants located in uMgungundlovu and eThekwini (Pillay *et al*., 2021). The study produced useful results for quantifying spatial and temporal variation in COVID-19 cases. However, the study was not without limitations as its spatial coverage (up to subdistrict and based on the wastewater plants) was coarse and thus, was unable to show granular level variations in COVID-19 cases.

From the literature reviewed above, it is noticeable that advancements in technology have revolutionised epidemiology and have made GIS analysis a critical part of epidemiological studies, allowing for disease spatial and temporal pattern identification and providing a unique view of pandemics that is otherwise unobtainable through other methods, such as purely statistical methods (Krieger, 2003; Rytkönen, 2004). It can also be said that geostatistical analysis and GIS-based models have proven to be valuable in surveillance, management and prevention of epidemics, including understanding and curbing COVID-19 (Clarke *et al*., 1996; Saffary *et al*., 2020). Also, it is noticeable that spatial modelling of COVID-19 within the South African context has been limited and has either used non-robust modelling methods or does not consider the granular level variations in the spatial distribution of COVID-19, leading to results that are misleading when planning COVID-19 response strategies. Subsequently, a gap still exists in studying micro-level COVID-19 incidence using advanced geospatial models that consider micro-level COVID-19 distribution within the South African context.

CHAPTER 3

Methods and materials

3.1. Study area

The research area for this study is the eThekwini Metro Municipality. The Municipality is situated on South Africa's east coast within the KwaZulu-Natal province (**Figure 1**) (Mbambo and Agbola, 2020). The Metro spans a total area of 2 555 km^2 and lies at 31° 01' E (longitude), and 29° 52' S (latitude) (eThekwini Municipality, 2020). Additionally, eThekwini Metropolitan is nestled between iLembe, Ugu, uMgungundlovu and the Indian Ocean in the north, south, west and east, correspondingly. (Jagarnath *et al*., 2020). Moreover, the research area is located within the humid subtropical region which is characterised by humid, hot summers with mean temperatures of around 24°C and mild winters with mean temperatures of about 17°C (Jagarnath *et al*., 2020). In addition to the above, the eThekwini Metro is characterised by an approximate population of 3.9 million people with an accompanying density of approximately 1448 people per square kilometre, houses the 3rd largest metropolitan in South Africa and is home to some of the major townships and informal areas in the country, including KwaMashu and Umlazi (Mbambo and Agbola, 2020; Pillay *et al*., 2021).

Moreover, eThekwini contributes significantly to the economy of South Africa through tourism, the manufacturing industry, and the sugar refining industry which makes up a great portion of South Africa's successful sugar production industry, which produces about 19.9 million tons of sugar (Reddy and Govender, 2019). It is noticeable that a large part of the population within the eThekwini Metro is poverty-stricken, with about one million people living well below the poverty threshold while the lack of access to resources including suitable water and sanitation is experienced by approximately 0.9% of the eThekwini population (eThekwini Municipality, 2020). Moreover, with approximately 5 655 fatalities and 346 987 COVID-19 cases at the time of the study, the study area has been a significant hotspot of the coronavirus (KZN DoH, 2022). Additionally, the chosen study area was one of the key areas targeted for COVID-19 vaccinations since the start of the vaccination programme on the 17th of February 2021 (SAcoronavirus, 2021). Consequently, eThekwini was a fitting study area to explore COVID-19 distribution at granular levels as it had most of the potential determining

factors of COVID-19 incidence, such as overcrowding, vaccine uptake and poverty, and has been severely impacted by the epidemic, and requires accurate pandemic control strategies through geospatial modelling. The availability of granular level COVID-19 data is another factor that made the eThekwini Metro a suitable study area.

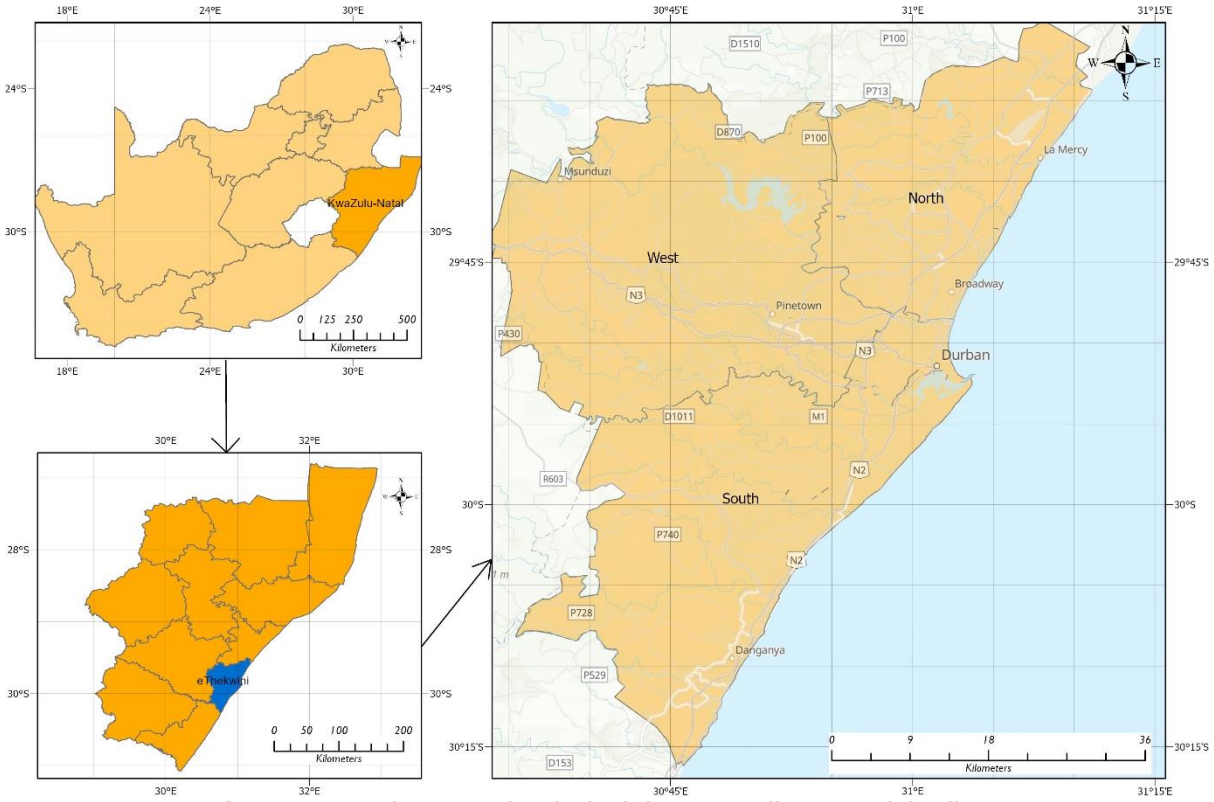


Figure 1. Study area of eThekwini Metropolitan Municipality.

3.2. Data acquisition and pre-processing

Research GIS data included shapefiles for boundaries, confirmed COVID-19 cases and other spatially enabled datasets which were freely acquired from various sources including the Municipal Demarcation Board (MDB), the District Health Information System (DHIS), the WorldPop repository, eThekwini municipality, and NDoH. The pre-processing of the GIS datasets involved cleaning up the data spreadsheets to eliminate confidential data and empty values and aggregating the data in a geodatabase.

Data about the number of settlement level confirmed coronavirus cases between the 2nd of May to the 11th of September 2021, which is the period spanning the severe 3rd wave of the pandemic characterised by an increase in vaccination coverage, was acquired in a digital Excel format from the municipality of eThekwini and cleaned to eliminate any confidential data and empty values. As part of additional data

processing, 34 431 micro-level coronavirus cases were geocoded using an excel VBA script and Google API and subsequently, polished through the elimination of any cases falling outside of the eThekwini Metro to lower noise. Following this procedure, there were 33,850 cases which had correct coordinates. Notably, the COVID-19 cases data was not split into training and validation sets as Moran's *I* technique sufficiently quantified spatial autocorrelation without the need for a validation dataset and a predictive model was not required (Dutta *et al*., 2021).

Additional information on the possible variables or factors affecting COVID-19 spread was sourced through the review of the literature and resulted in the identification of ten COVID-19 incidence determining factors which can be categorized into social, demographic, environmental, comorbidity, health security, and economic factors (**Table 1**). Data on the social predictor of COVID-19 incidence in the form of population density was sourced from the GeoTerra Image (GTI) 2019 population estimates geodatabase and from census data coming from Statistics SA 2011 as the latest census data had not been released during the completion of the study.

Additionally, data about COVID-19 behavioural responses and the impacts thereof on COVID-19 spread was collected through Statistics SA's coronavirus vulnerability information, which was developed to evaluate localised COVID-19 risk factors at the small area level (Statistics South Africa, 2022). The COVID-19 vulnerability indices were aggregated by settlements in a shapefile and accounted for various behaviours influencing COVID-19 spread including social distancing, mask-wearing, and sanitisation.

Moreover, data on the proportion of the population living under the poverty threshold was obtained from the 2011 census data from Statistics SA while demographic data relating to age distribution and gender were obtained from the 2019 GTI estimates, which were also accessible in a geodatabase. Data on vaccination coverage at a settlement level was obtained from the NDoH along with data on health facilities. This data was obtained in Excel format and cleaned to align with the requirements of the study. Data on the economic determinant of COVID-19 spread in the form of the wellness score was also obtained from the GTI estimates while data on the prevalence of the tuberculosis was obtained from DHIS in Excel format and cleaned to align with settlement boundaries. Additionally, data on air pollution concentration from 1998 to

2016 was obtained from the Living Atlas as the latest available data on the local level concentrations of particulate matter. The data is readily available in a GIS format. Acquired data was aligned with the study period (May to September 2021) where necessary. Noticeably, when conducting geospatial analysis and modelling, the reliant variable was assumed to be the data on coronavirus incidence while the predictor variables were taken as the independent variables.

Table 1. List and details of COVID-19 spread explanatory variables which were selected as factors that have a significant impact on COVID-19 incidence at local levels and were used as independent variables for the spatial models.

Category	Predictor	Indicator	Scale	Data Source	Data Format
Environment	Air pollution	Particulate matter concentrations (μg/m^3)	50 km hexagon grids	Living Atlas	Shapefile
Social	Overcrowding	Population density (people/km^2)	Settlement level	Statistics SA 2011 & GTI estimates 2019	Shapefile
	COVID-19 behavioural response	COVID-19 Vulnerability Index	Enumeration area	Statistics SA 2011	Shapefile
Economic	Income	Wellness score	Enumeration area	GTI Estimates 2019	Shapefile
	Poverty	Population living below the poverty line (%)	1×1 km square grids	Statistics SA 2011	Shapefile
Demographic	Age distribution	Primary dominant age group	Enumeration area	GTI Estimates 2019	Shapefile
	Gender	Primary dominant gender	Enumeration area	GTI Estimates 2019	Shapefile
Comorbidity	Tuberculosis prevalence	Tuberculosis cases per settlement	Health facility level	DHIS	Excel File
Health security	Health facility density	Density of health facilities per settlement	Settlement level	NDoH	Excel File
	Vaccine uptake	Vaccination coverage (%)	Settlement level	NDoH	Excel File

3.3. Data analysis

The sourced data was analysed through advanced spatial modelling including spatial autocorrelation. The analysis of data entailed mapping coronavirus incidence or cases at the micro-levels across the eThekwini Metro and employing spatial autocorrelation (Moran's *I*) to explore their spatial spread; the spatial autocorrelation was conducted using *ArcGIS Pro* and *GeoDa*, which both perform adequately for geostatistical analysis and modelling that (Dutta *et al*., 2021; Esri, 2022b). Moran's *I,* which is used for spatial autocorrelation, is the regression analysis method which was employed for the analysis and modelling of the incidence of COVID-19 and spatial patterns, and that has been used comprehensively in other studies aiming to spatially model COVID-19 and has produced valuable outcomes (Mollalo *et al*., 2020; Saffary *et al*., 2020; Castro *et al*., 2021; Hassaan *et al*., 2021; Sangkham *et al*., 2021; Dutta *et al*., 2021). Noticeably, spatial autocorrelation quantifies the degree of relation between observations within a locality- checks for the presence of spatial patterns- and in most cases commonly uses Moran's index which is computed through the below formula:

$$I = \frac{1}{s^2} \frac{\sum_i \sum_j (y_i - \bar{y})(y_j - \bar{y})}{\sum_i \sum_j w_{ij}}$$

where the elements in the spatial weight are represented by *i* and *j* and the likeness between these components is computed by multiplying the change in y_i and y_j with the overall average divided by s^2, which is the variance of the model (Shariati *et al*., 2020; Hassaan *et al*., 2021; Dutta *et al*., 2021). Additionally, spatial autocorrelation is defined as a statistical quantification of patterns between objects within a space or spatial association and is characterised by a null hypothesis stating that spatial features have values that are not spatially associated and an alternative hypothesis stating that the spatial features have spatially associated values (Waldhör, 1996; Esri, 2022b). The outcome of the above formula is a Moran's *I* statistic, which shows the spatial distribution pattern of the data and highlights low and high-risk clusters which can be processed for further analysis (Dutta *et al*., 2021). Noticeably, Moran's *I* scatterplots along with optional LISA significant maps are used for visualising the above-mentioned significant clusters while Local Indicators of Spatial Association maps are utilised to visualise the clusters (Hassaan *et al*., 2021; Dutta *et al*., 2021). The outcomes of the spatial autocorrelation indicated the settlement level pattern of COVID-19 incidence and highlighted outliers (both statistical and spatial) which needed to be eliminated to

lessen noise. Additionally, the correlation between significant coronavirus incidences and significant determinants was illustrated through bivariate LISA maps of clusters and significance (Dutta *et al*., 2021).

The geographically weighted regression (GWR) model was employed as a local approach for exploring spatially varying associations between a group of independent variables and a dependent factor (Brunsdon *et al*., 1996). Notably, the GWR approach is useful for modelling data that has spatial heterogeneity as it estimates separate regression models for individual locations in the dataset (Brunsdon *et al*., 1996; Dutta *et al*., 2021). The GWR model has the formula:

$$y_i = \beta_{i0} + \sum_{j-1}^{m} \beta_{ij} X_{ij} + \varepsilon_i,\ i = 1,\ 2,\ \ldots,\ n$$

where *i* is the geographic point where coefficients are projected, y_i is the reliant term, β_{i0} denotes the intercept, β_{ij} is the regression factor, while X_{ij}, is the value of the j_{th} determinant and the arbitrary variable for error is ε_i.

In addition to the above, the spatial association amongst the determinants and the occurrence of COVID-19 was further modelled through MGWR which is an extension of the GWR model which models and estimates spatial associations between factors at multiple local scales (Hassaan *et al*., 2021). Noticeably, MGWR analyses local level spatial associations and accounts for spatial diversity through estimating coefficients at multiple bandwidths, making it a fitting model for micro-level analysis of COVID-19 incidence (Dutta *et al*., 2021). The formula for MGWR is:

$$y_i = \sum_{j-0}^{m} \beta_{bwj} X_{ij} + \varepsilon_i,\ i = 1,\ 2,\ \ldots,\ n$$

where the calibration bandwidth is denoted by β_{bwj}, y_i is the dependent variable in location *i*, ε_i is a random variable for error, and the remainder of the factors are the same as those of the GWR model (Dutta *et al*., 2021).

OLS regression was also utilised to investigate the applicability of advanced spatial models, including spatial autocorrelation techniques, relative to the standard regression techniques that are not specifically designed to model spatial

dependencies. Notably, OLS is a type of linear (either simple or multiple) regression for explaining associations between dependent and independent variables (Esri, 2022a). OLS is represented by the equation:

$$y_i = \beta_0 + x_i\beta + \varepsilon_i$$

where the dependent variable is y_i, the intercept is β_0, β is the modelling coefficients vector, x_i is the vector of the determinants, and ε_i is the random variable for error. Notably, the coefficient β for each independent factor being equal to 0 is the null hypothesis for the OLS regression and β not being equal to 0 is the alternative hypothesis (Esri, 2022a).

The spatial error model (SEM) and the spatial lag model (SLM) were also utilised to investigate how global spatial models fare to GWR and MGWR which are local spatial models. Noticeably, SLM models the relation between the predictor and the dependent variable through adding a spatial lag term which regresses the dependent factor based on neighbouring values and the values of the dependent variable. SLM is represented by the formula:

$$y_i = \beta_0 + x_i\beta + \rho W_i y_i + \varepsilon_i$$

where ρ is a term for spatial lag, the spatial weights vector is denoted by W_i and the rest of the variables are the same as the parameters of the OLS model but have a disintegrated error parameter (Dutta *et al*., 2021). Similarly, the SEM models the relationship between the predictor and the dependent variable through adding a spatially structured error term to account for the residuals' spatial autocorrelation (Mollalo *et al*., 2021; Dutta *et al*., 2021). SEM is denoted by:

$$y_i = \beta_0 + x_i\beta + \lambda W_i\xi_i + \varepsilon_i$$

where the spatial module of the error is denoted by ξi, λ denotes the association degree between modules, and ε_i is the spatially unrelated error parameter (Mollalo *et al*., 2021; Dutta *et al*., 2021).

3.4. Model validation

As noted above, variables that were insignificant, whether spatially or statistically, were discarded. Thereafter, the multicollinearity of data amongst the predictor variables was addressed before spatial autocorrelation analysis (Brunsdon *et al*., 1996; Waldhör, 1996; Kindi *et al*., 2021). The above was achieved by calculating the variance inflation factors (VIF) through OLS regression, with redundancy being illustrated by VIF values higher than 7.5 (Ahasan *et al*., 2022; Esri, 2022; Dutta *et al*., 2021). The global models and spatial autocorrelation models were computed in *ArcGIS Pro* 3.2 and *GeoDa* 1.2. The performance of Moran's *I* and the global models in modelling the spatial relation between the dependent variable (incidence of COVID-19) and the predictors was validated through computing R^2 and Akaike Information Criterion (AIC) values.

Additionally, the models used have several assumptions which have an impact on the accuracy and reliability of the results. The SEM model used to compare the performance of the global models with that of local models assumes that errors are spatially corrected and normally distributed, there is no multicollinearity, and the relationship between the dependent and independent variables is linear (Zhang and Song, 2024). Additionally, the SEM model has the null hypothesis that the spatial autocorrelation coefficient is equal to zero, indicating the absence of spatial autocorrelation (Ahasan et al., 2022; Dutta et al., 2021; Zhang and Song, 2024). Subsequently, SEM results that show no multicollinearity and constant variance across observations indicate that the model does not overestimate or understate variability and produces reliable results (Dutta *et al*., 2021). Similarly, the SLM model also assumes that spatial autocorrelation exists in the dependent variable, there is no multicollinearity in the dependent variable, and that there must be linearity between the explanatory and dependent variables (Akinwumiju *et al*., 2022; Zhang and Song, 2024). The null hypothesis of the SLM model is that the spatial lag parameter is equal to zero and implies that the model is suitable for capturing spatial dependence if the hypothesis is rejected (Dutta et al., 2021).

It is also noticeable that the OLS model has the assumptions that the relationship between the dependent and independent variables is linear and that there is no multicollinearity (Mollalo *et al.,* 2020). Additionally, the normal distribution of errors and the independence of observations from each other are also additional assumptions of

the model. Notably, the null hypothesis of the model is that independent variables coefficients are equal to zero and rejecting the null hypothesis suggests that the explanatory variables have a statistically significant effect on the dependent variable, thus making the results reliable and relevant (Mollalo *et al.,* 2020). In addition to the above, the GWR and MGWR models both assume that there is linearity between the dependent and independent variables, the there is no multicollinearity, there is constant variance across space and that relationships between variables are spatially correlated (Mahara *et al.,* 2018; Irandoost *et al.,* 2023; Thammaboribal *et al.,* 2024). Similar to the other models, the null hypotheses for the GWR and MGWR models imply that the coefficient is constant across all locations, and that global models, such as OLS, are less suitable for modelling local-level spatial variability and geographically weighted approaches should be utilised (Franch-Pardo *et al.*, 2020).

CHAPTER 4

Results

4.1. Mapping the spatial spread of COVID-19 incidence

The spread of confirmed COVID-19 cases in settlements within the eThekwini Metro was mapped (**Figure 2**). It is noticeable that during the third wave, most settlements in the western, northern, and central parts had cases ranging between 100 to 500 while lower cases were noticeable in the southern parts of the Metro. Additionally, the highest sum of confirmed COVID-19 cases was noted in Berea and Bluff which are settlements on the central coast of the Metro. In addition to the above, spatial distribution results illustrated that a high number of COVID-19 cases was visible around the central and northern parts of the Metro (**Figure 2**). Low cases were notable around the southern parts of Adams Mission and Umbumbulu settlements.

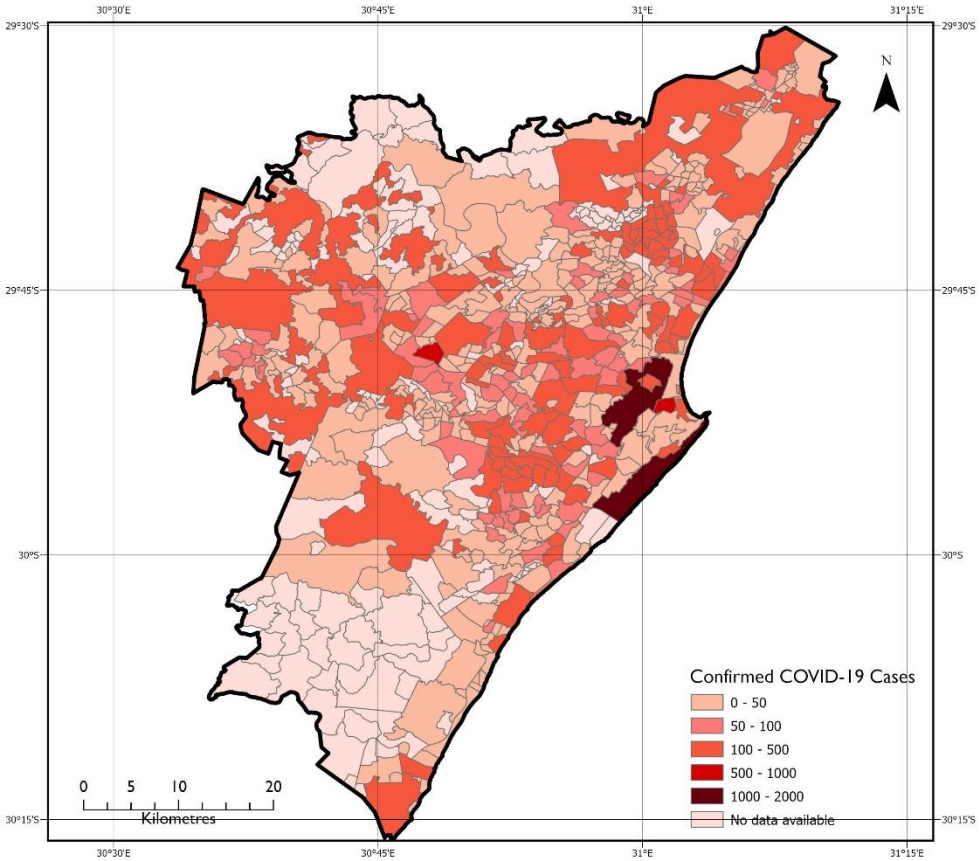


Figure 2. Map of the spatial distribution of confirmed COVID-19 cases in the eThekwini Metro during the third wave (May 02 to September 11, 2021).

Following the elimination of spatial and statistical outliers, Moran's *I* result showed a positive correlation in COVID-19 incidence in eThekwini (**Figure 3**). The COVID-19 incidence Moran's *I* was relatively small with a value of 0.14, while a higher z-score of 6.60 was notable and accompanied by a small p-value (0.0) (C). Additionally, Moran's I result illustrates that COVID-19 spread within eThekwini had a clustered pattern. The LISA cluster map (A) illustrates that Low-Low clusters (blue) are dominant in the north-western and southern regions of the study area while Low-High (lavender) and High-High clusters (red) are notable in the eastern regions of the area, and High-Low clusters (peach) are dominant on the western and northern regions of the study area. Additionally, it is visible that both the Low-Low and High-High clusters carried statistical weight as seen from p-values (0.001 and 0.01, respectively) (B).

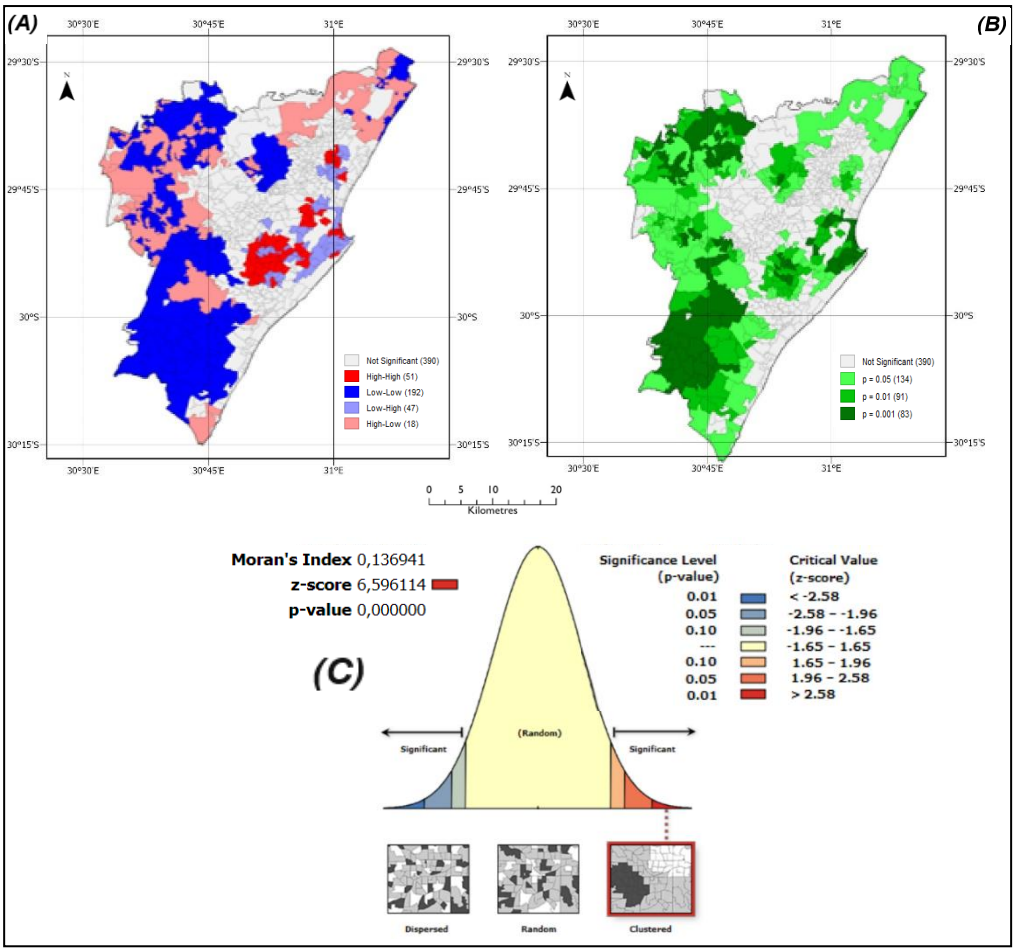


Figure 3. LISA cluster map for COVID-19 cases for eThekwini settlements (A) along with the significant map of clusters shown with the associated p-values (B) and the spatial autocorrelation report showing Moran's Index, z-score, and p-value for COVID-19 incidence (C).

4.2. Comparing global and local models for COVID-19 incidence analysis

Following the OLS analysis there were only three statistically significant explanatory variables with p-values (probability) less than 0.05 (**Table 2**). These are wellness score, population density, and TB cases. The rest of the explanatory variables were statistically insignificant as their p-values were greater than 0.05. Additionally, a positive association between COVID-19 incidence and population density (coefficient = 0.000366), economic wellness (coefficient = 89.70) and the number of TB cases (coefficient = 0.43) is noticeable, while a decreasing correlation between COVID-19 incidence and primary dominant gender, COVID-19 vulnerability index, health facility density and particulate matter concentrations is also noticeable. Moreover, the VIFs for all factors were comparatively small and lower than the cut-off of 7.5, which highlighted the absence of local multicollinearity (Esri, 2022a). Additionally, the R^2 value for the OLS model was 0.398 which indicates that 39.8% of the variation is explained by the model. Notably, this value illustrates the presence of a moderate effect on the spread of the virus.

Table 2. Ordinary Least Squares results indicating the overall goodness of fit (local R^2 and overall adjusted R^2), the significance of the coefficients (β, t-statistic and probability), the level of multicollinearity (VIF) and the overall significance of the model (AIC).

Variable	Coefficient (β)	t-Statistic	Probability	VIF
Intercept	-169.37	-0.62	0.54	-
Population density	0.000366	6.23	0.00	1.34
Economic wellness score	89.70	2.44	0.02	2.41
EVDS vaccination coverage	0.75	0.73	0.47	2.36
Primary dominant gender	-4.71	-0.12	0.91	1.15
Number of TB cases	0.43	2.35	0.02	1.33
Primary dominant age group	16.78	0.48	0.64	1.91
COVID-19 vulnerability index	-39.99	-0.71	0.48	2.21
Population living below poverty line (%)	1.88	0.93	0.361	2.58
Health facility density	-174.14	-0.47	0.64	2.28
Particulate matter concentrations ($\mu g/m^3$)	-3.62	-0.31	0.76	2.54
Local R^2 = 0.398				
Overall adjusted R^2 = 0.387				
AIC = 6665.11				

The SEM and SLM models exhibited improved performance in comparison to the OLS model. The above is corroborated by the summary statistics presented below (**Table 3**). Similar to the OLS results, wellness score, population density and the number of TB cases are the statistically significant explanatory variables of COVID-19 spread according to the SEM and SLM results. Additionally, the moderately high Lambda values of the SEM model varying from 0.03 to 0.57 along with the relatively high lag coefficient on the SLM model indicate the existence of spatial dependence between the statistically significant variables and COVID-19 cases. Moreover, the increasing local R^2 from 0.398 in the OLS model to 0.399 in the SEM model along with the decreasing AIC values (6665.11 in the OLS model to 6662.67 in the SEM model) indicate the improved performance of the SEM and SLM models in explaining variance in COVID-19 incidence and that the models adequately account for the variables that are potential determinants of COVID-19 spread.

Table 3. SLM and SEM results showing the significance of the coefficients (coefficient, standard error, Z-score, and probability), goodness of fit of the models (local R^2) and the overall performance of the models in modelling COVID-19 incidence (AIC).

Variable	**Coefficient**		**Standard Error**		**Z-score**		**Probability**	
	SEM	SLM	SEM	SLM	SEM	SLM	SEM	SLM
Intercept	76.38	81.13	52.65	52.42	1.45	1.55	0.15	0.12
Population density	0.00025	0.00024	0.000025	0.000025	10.28	10.04	0.00	0.00
Economic wellness score	14.81	14.53	6.01	5.99	2.46	2.43	0.01	0.02
Primary dominant gender	-9.06	-9.78	8.28	8.27	-1.09	-1.18	0.27	0.24
EVDS vaccination coverage	0.124	0.12	0.08	0.08	1.55	1.61	0.12	0.11
Number of TB cases	0.55	0.56	0.08	0.08	6.54	6.65	0.00	0.00
Primary dominant age group	7.06	7.28	6.50	6.49	1.09	1.12	0.28	0.26
COVID-19 vulnerability index	-13.10	-13.21	9.24	9.23	-1.42	-1.43	0.16	0.15
Population living below poverty line (%)	-0.56	-0.59	0.42	0.41	0.18	-1.44	0.17	0.15

Health facility density	2.06	14.92	86.53	85.57	0.02	0.17	0.98	0.86
Particulate matter concentrations (µg/m^3)	-2.73	-2.90	2.62	2.57	- 1.0442	-1.13	0.30	0.26
LAMBDA	0.03	-	0.06	-	0.57	-	0.57	-
W_Covid Cases	-	-0.009	-	0.044	-	-0.207	-	0.836

SEM Local R^2 = 0.399
SLM Local R^2 = 0.398

SEM AIC = 6662.67

SLM AIC = 6663.19

Compared to the OLS, SEM and SML results, the GWR and MGWR results suggest that the local level spatial models exhibited improved performance and resulted in more statistically significant results than the global models (**Table 4** and **Table 5**). The above is evidenced in the local R^2 values (0.716 for GWR and 0.725 for MGWR) which were significantly higher than those of the global models and indicate that the GWR model explained 71.6% of the coronavirus spread variance while MGWR explained 72.5% of the variance. Consequently, the GWR and MGWR models explained almost 33% more of the variance in COVID-19 incidence compared to the global models (OLS, SEM and SLM). Moreover, the comparatively lower AIC values (199.742 and 193.392 for GWR Sand MGWR, respectively) indicate that these models had better explanatory power and overall fit for the data.

Table 4. Geographically Weighted Regression results indicating the goodness of fit of the model (R^2 and adjusted R^2), spatial variations in the relationships between the determinants and COVID-19 cases (minimum, maximum, mean, and standard deviation values) and the overall performance of the GWR model (AIC).

Variable	Minimum	Maximum	Mean	Standard deviation
Intercept	-0.11	0.11	0.03	0.05
Population density	0.05	0.81	0.58	0.17
Economic wellness score	0.05	0.54	0.27	0.10
Primary dominant gender	- 0.04	0.03	-0.00	0.01
EVDS vaccination coverage	-0.05	0.20	0.07	0.07
Number of TB cases	0.04	0.22	0.16	0.04
Primary dominant age group	-0.10	0.53	0.11	0.14
COVID-19 vulnerability index	-0.31	0.05	-0.13	0.10
Population living below poverty line (%)	-0.31	0.78	0.13	0.17
Health facility density	-0.24	0.01	-0.09	0.07
Particulate matter concentrations (μg/m^3)	-0.26	0.11	-0.04	0.11
Local R^2 = 0.716				
Overall adjusted R^2 = 0.608				
AIC = 199.742				

Table 5. Multiscale Geographically Weighted Regression results indicating the goodness of fit of the model (R^2 and adjusted R^2), spatial variations in the relationships between the determinants and COVID-19 cases (minimum, maximum, mean, and standard deviation values) and the overall performance of the GWR model (AIC).

Variable	Minimum	Maximum	Mean	Standard deviation
Intercept	- 0.06	-0.04	-0.05	0.00
Population density	0.08	0.93	0.44	0.22
Economic wellness score	0.20	0.42	0.31	0.05
Primary dominant gender	- 0.01	0.02	0.01	0.01
EVDS vaccination coverage	0.04	0.10	0.07	0.01
Number of TB cases	0.13	0.16	0.15	0.01
Primary dominant age group	- 0.08	0.63	0.16	0.12
COVID-19 vulnerability index	- 0.27	-0.02	-0.13	0.07
Population living below poverty line (%)	0.07	0.55	0.22	0.12
Health facility density	- 0.06	0.02	-0.02	0.02
Particulate matter concentrations (μg/m^3)	- 0.05	0.01	-0.01	0.01
Local R^2 = 0.725				
Overall adjusted R^2 = 0.629				
AIC = 193.392				

4.3. Spatial dependence of COVID-19 incidence

The bivariate LISA results for COVID-19 incidence and population density are presented below (**Figure 4**). It is noticeable that population density and COVID-19 incidence have a moderately high positive spatial association (Moran's *I* = 0.26 and p-value = 0.00) (C). The LISA cluster map highlights the similarities in the spatial spread of coronavirus cases and population density as there are several Low-Low clusters in the western and southern regions of the Metro, and several High-High clusters in the central and north-eastern regions of the research site covering Durban Central, KwaMashu and Umlazi (A). Additionally, the Low-Low and High-High clusters are all statistically significant as they have p-values lower than or equal to 0.05 (B). Thus, population density is positively spatially associated with coronavirus incidence.

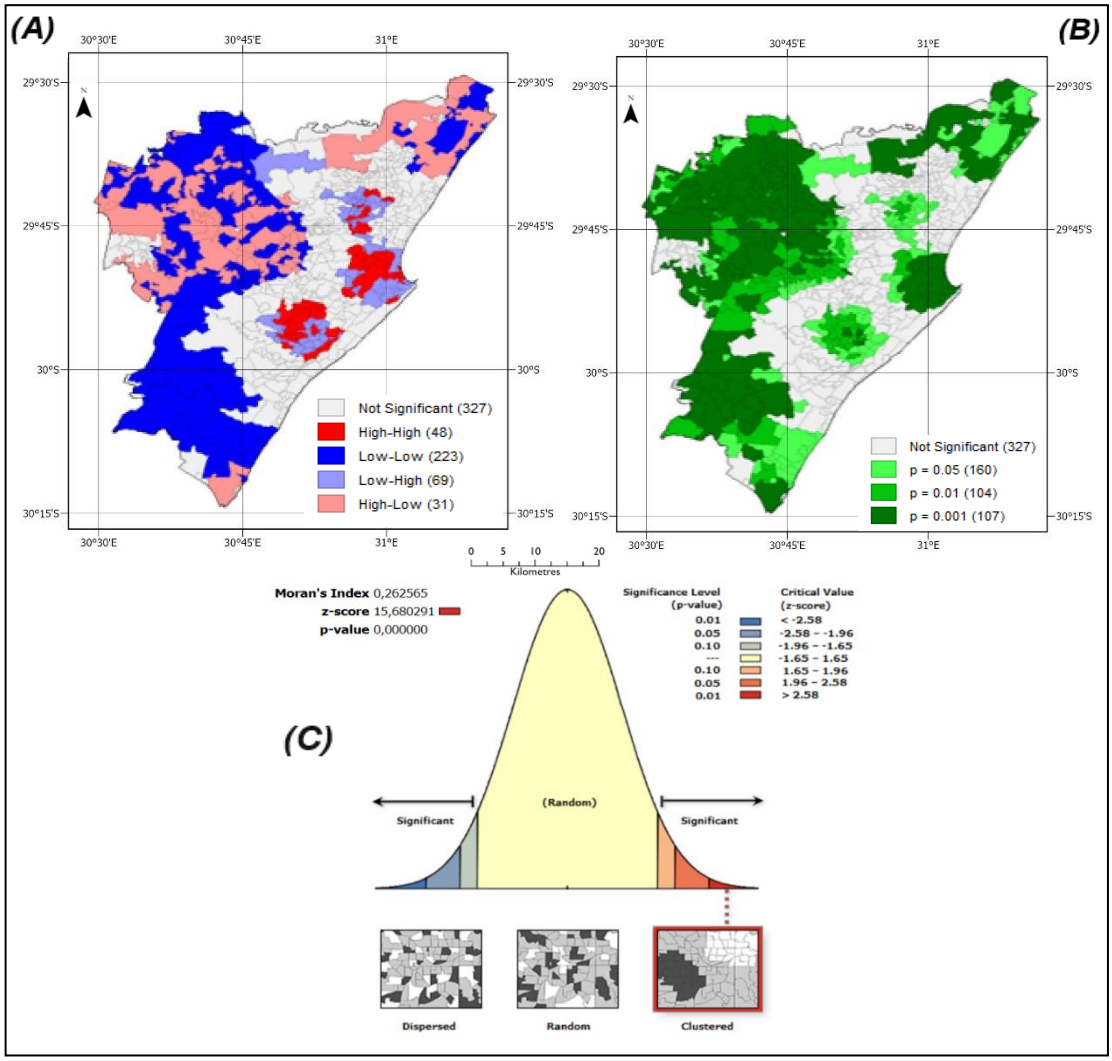


Figure 4. Bivariate LISA cluster map for COVID-19 cases for eThekwini settlements and population density (A) along with the significant map of clusters shown with the associated p-values (B) and the spatial autocorrelation report showing Moran's Index, z-score, and p-value for population density and COVID-19 incidence (C).

The bivariate LISA results for COVID-19 incidence and economic wellness score are presented below (**Figure 5**). It is noticeable that economic wellness and COVID-19 incidence at a settlement level have a relatively high positive spatial autocorrelation as shown by the value of Moran's *I* (0.30) and the p-value of 0.00 (C). The Low-Low and High-High clusters between COVID-19 incidences and economic wellness exhibit similar patterns (A). Noticeably, Low-Low clusters are more dominant in the north-western and southern regions of the eThekwini Metro while High-High clusters are more noticeable in the central and eastern portions of the Metro. Additionally, the High-High clusters mostly had p-values of 0.001 while p-values of 0.01 and 0.05 were more notable in the Low-Low clusters. Consequently, COVID-19 incidence and economic wellness are positively spatially correlated (B).

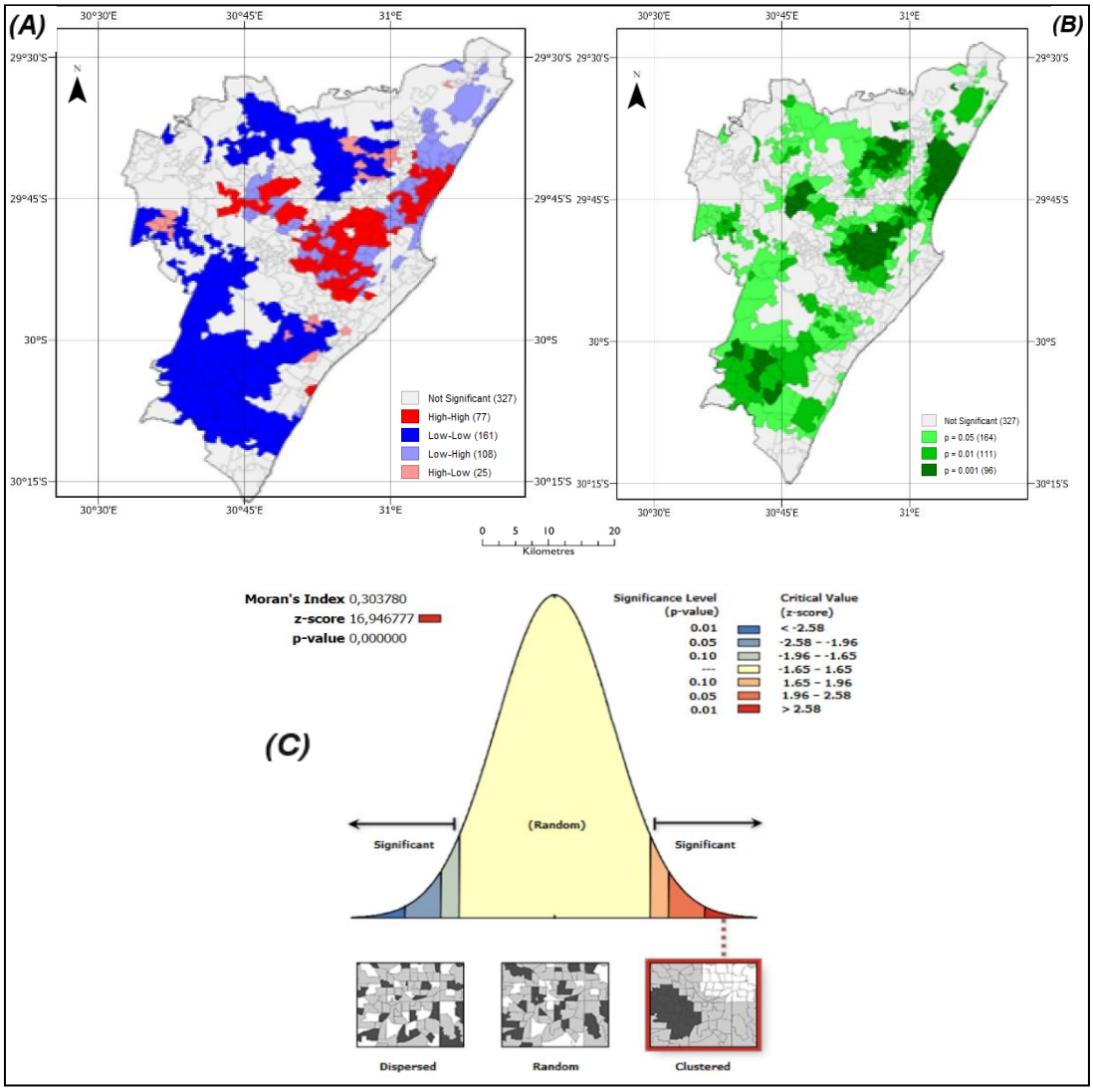


Figure 5. Bivariate LISA cluster map for COVID-19 cases for eThekwini settlements and economic wellness score (A) along with the significant map of clusters shown with the associated p-values (B) and the spatial autocorrelation report showing Moran's Index, z-score, and p-value for economic wellness and COVID-19 incidence (C).

Similarly, spatial dependence exists between TB prevalence and COVID-19 incidence (**Figure 6**). The above is corroborated by the spatial autocorrelation report (C) which shows that TB prevalence and COVID-19 incidence have a clustered pattern in eThekwini (Moran's *I* of 0.13) along with the LISA cluster map (A) which shows that there are a few High-High clusters around Durban Central, Berea and some parts of Umlazi, and High-Low clusters in the southern parts of the study site. Additionally, the significant map (B) illustrates that the clusters are statistically significant as the p-values are lower than 0.05. Thus, the relationship between the number of TB cases is spatially associated with COVID-19 incidence.

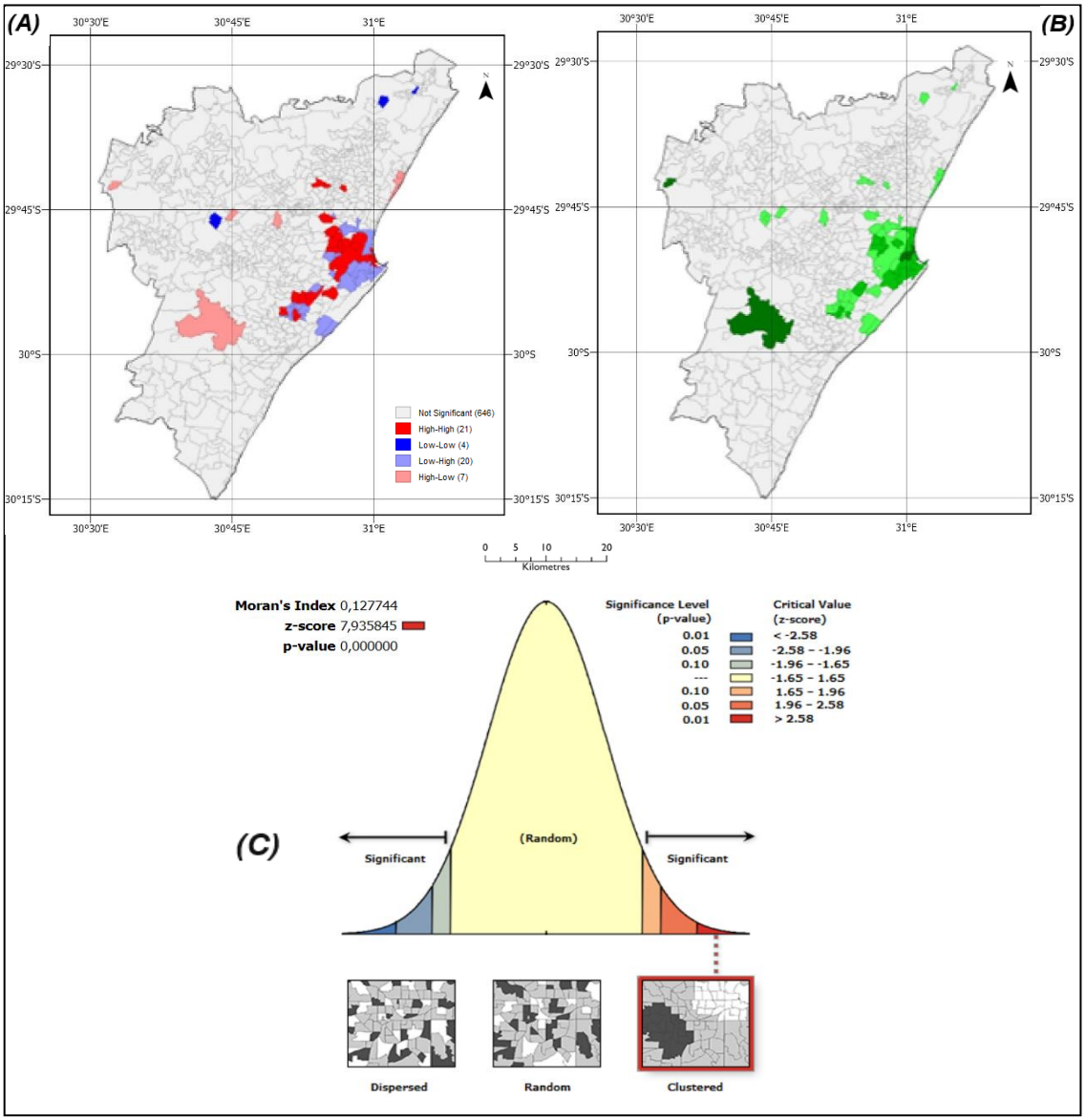


Figure 6. Bivariate LISA cluster map for COVID-19 cases for eThekwini settlements and number of TB cases (A) along with the significant map of clusters shown with the associated p-values (B) and the spatial autocorrelation report showing Moran's Index, z-score, and p-value for the number of TB cases and COVID-19 incidence (C).

4.4. Modelling spatial effects of determinants on COVID-19 incidence

The maps derived from the GWR and MGWR models for the three variables that were statistically significant (population density, economic wellness score and number of TB cases) are presented below (**Figure 7**, **Figure 8** and **Figure 9**).

The impact of population density in explaining COVID-19 incidence is shown below (**Figure 7**). Larger coefficients (0.000748 to 0.011336) are observable in the northern, southwestern, western, and central portions of the study area (**Figure 7**). Thus, the spatial association between population density and cases of COVID-19 in the above-mentioned areas is very strong and high population density is correlated with a high number of COVID-19 cases. Comparatively, some southern central and north-western settlements have small coefficients ranging from below 0 to 0.000286 and thus, indicate a strongly negative relationship between COVID-19 incidence and population density meaning that in these areas high population density is correlated with lower COVID-19 cases. Notably, the MGWR results are also comparable to the GWR results. However, the MGWR coefficients are slightly higher than those of the GWR with the highest values being 1.58 to 2.62) (B).

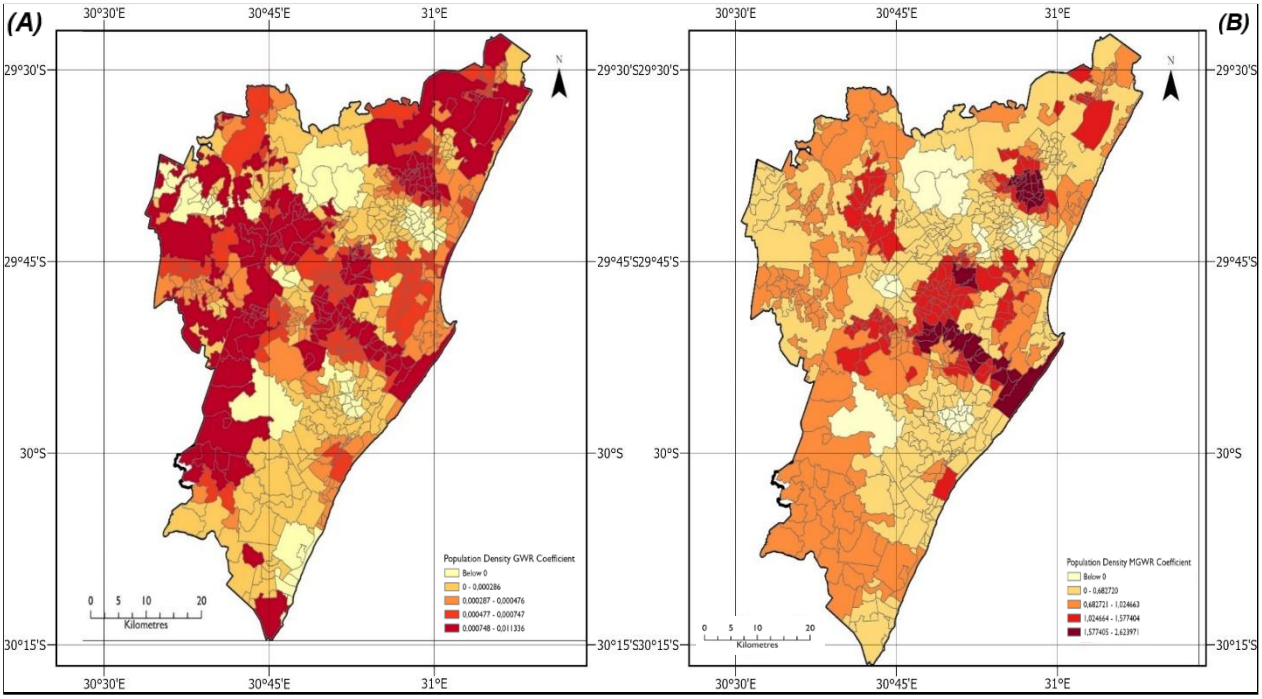


Figure 7. GWR (A) and MGWR (B) coefficient results for population density in describing COVID-19 incidence.

The effect of the economic wellness score in describing coronavirus incidence is shown below (**Figure 8**). Larger coefficients ranging from 0.22 to 0.97 are observable along the south-eastern and eastern portions of the coast extending into the upper central settlements. Thus, high economic wellness in the above-mentioned parts is correlated with a high number of COVID-19 cases. The above strong spatial association is also noticeable in the MGWR coefficients which range from 0 to 0.45 (B). Notably, the MGWR results have a more defined spatial pattern with the higher coefficients being on the coastal edge and slowly decreasing moving inland (B). Consequently, in the coastal settlements economic wellness has a stronger positive correlation with the incidence of COVID-19 in comparison to the inland settlements.

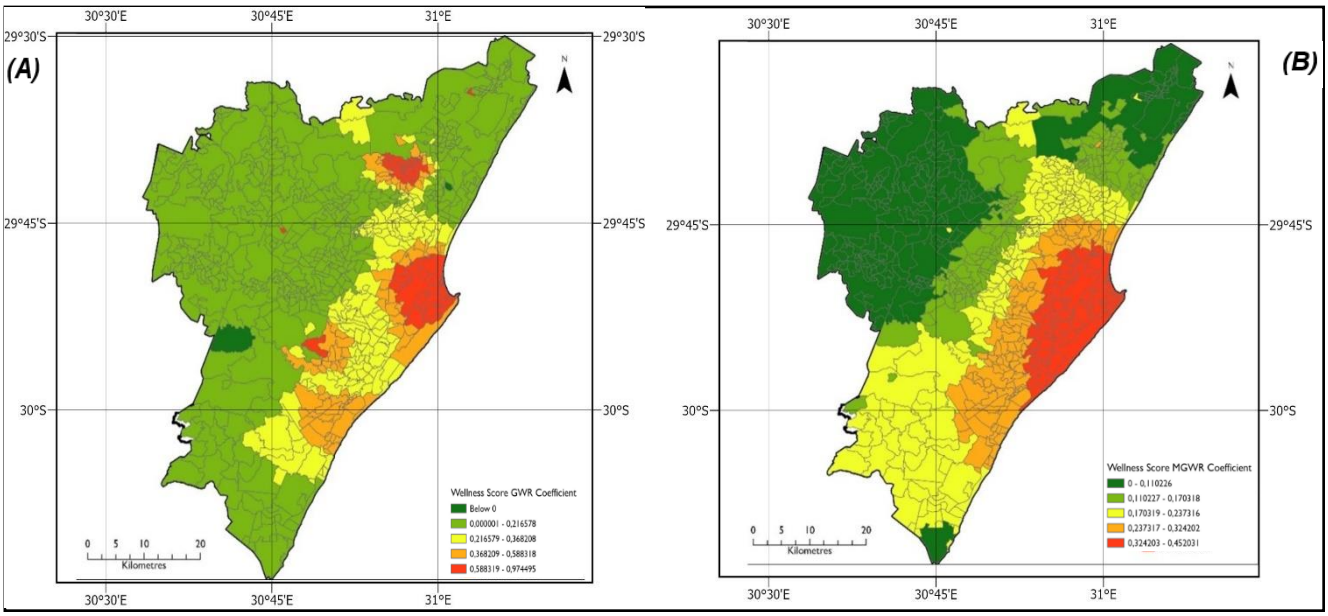


Figure 8. GWR (A) and MGWR (B) coefficient results for economic wellness score in describing COVID-19 incidence.

The GWR and MGWR coefficients for TB prevalence are presented below (**Figure 9**). Notably, there are several settlements which have no available data on both maps. Additionally, in the GWR map, the lower coefficients as little as 0.24 are more dominant in the western and northern parts of the Metro while higher coefficients are dominant in the eastern and southern parts thus, exhibiting strong positive relationships between COVID-19 incidence and the sum of TB cases (A). Notably, although the above-mentioned pattern is visible it is not consistent. A similar pattern is notable on the MGWR map where moderate coefficients are more dominant on the western and northern edges of the study area while much greater values are more notable on the eastern and southern parts and the lowest values are found in the northern central parts of the eThekwini Metro (B).

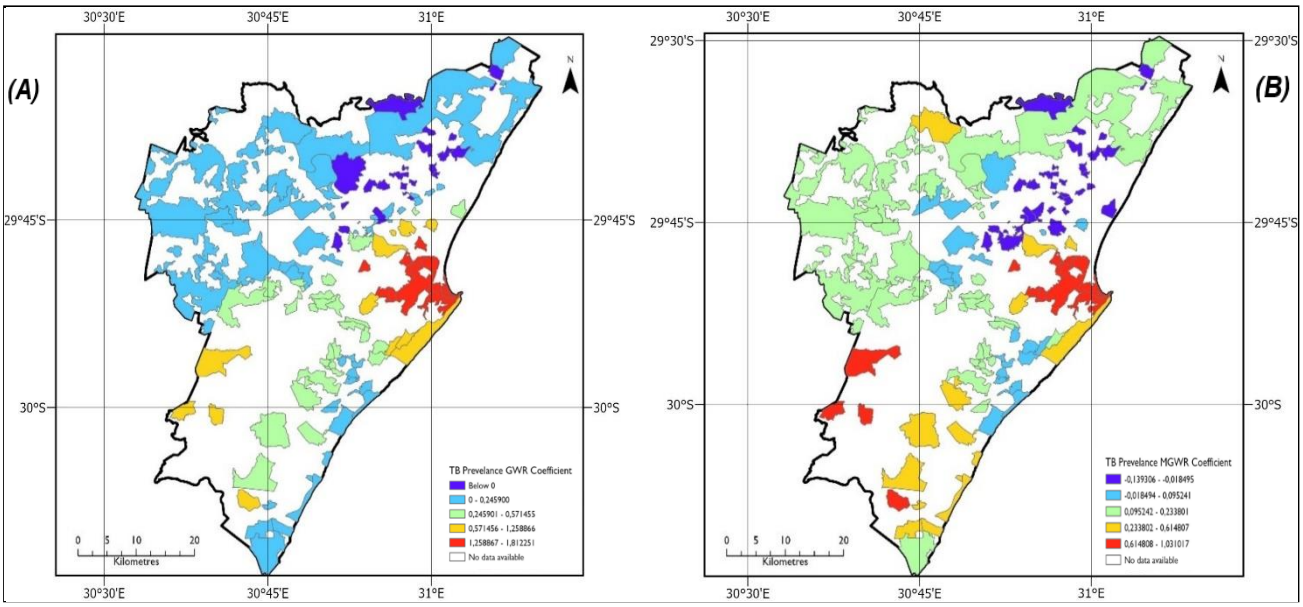


Figure 9. GWR (A) and MGWR (B) coefficient results for the number of TB cases in describing COVID-19 incidence.

The spatial distribution of the local R^2 values for the MGWR and GWR models for population density, economic wellness score and the number of TB cases is presented below (**Figure 10**). Lower R^2 values (0.13 to 0.45) are more dominant on the western, northern, and southern edges of the research site while larger values (0.55 to 0.72) are more noticeable on the central and eastern regions of the study area for GWR (A). A similar pattern is also noticeable on the MGWR with higher values varying from 0.58 to 0.72 being dominant on the central and eastern parts and lower values ranging from 0.01 to 0.46 being dominant on the northern, southern, and western areas of the Metro (B). The larger R^2 values highlight a better fit of the models in the settlements and thus, the models exhibited better explanatory power in the central and eastern regions of eThekwini.

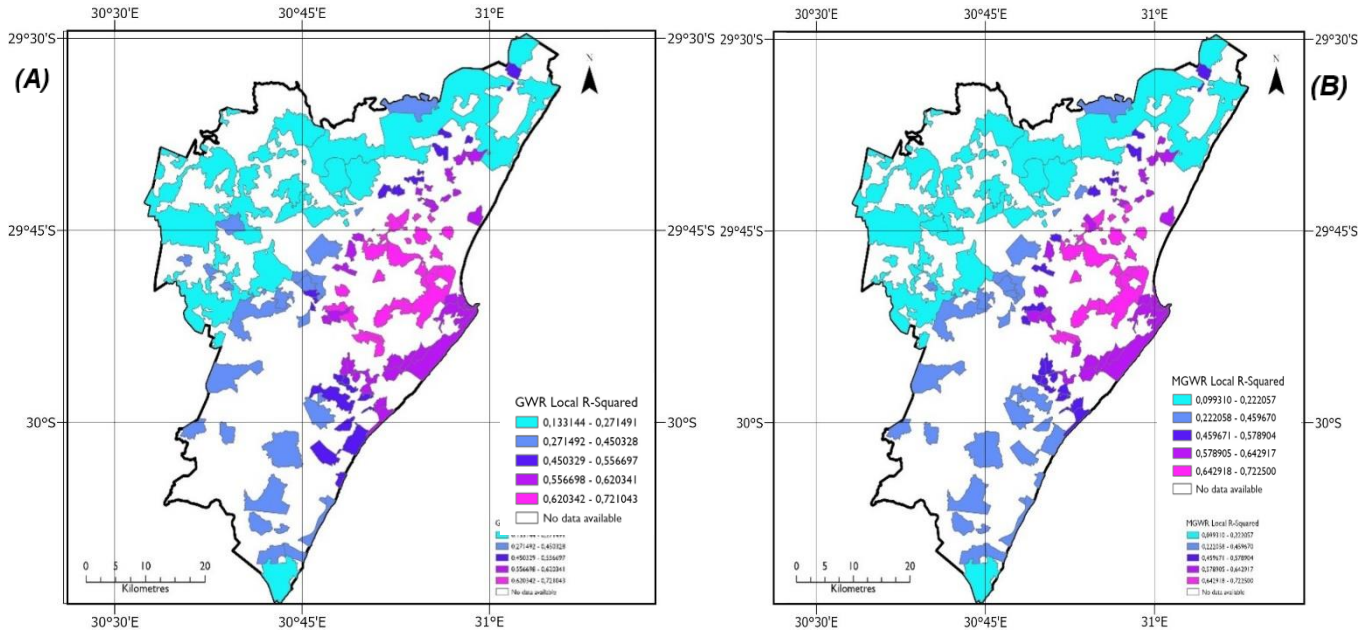


Figure 10. The spatial distribution of the local R^2 values for the GWR (A) and MGWR (B) models.

CHAPTER 5

Discussion

5.1. Spatial autocorrelation of COVID-19 incidence

In this study, spatial modelling in the form of spatial autocorrelation was utilised to map the granular level geographical distribution of COVID-19 over the 3rd wave (May 02 to September 11, 2021) in the eThekwini Metropolitan area and to model the dependence of COVID-19 incidence on micro-level factors. Ten factors were considered for the study as potential determining variables of the incidence of COVID-19 and were grouped into six categories including, environment, social, demographic, comorbidity, and health security. Following the assessment of the spatial outspread of the occurrence of COVID-19 in the eThekwini Metro through spatial autocorrelation, and the elimination of multicollinearity through OLS regression, significant contributors to the incidence of COVID-19 and the impacts thereof were identified through Moran's *I* significance maps.

From the spatial distribution results (**Figure 2**) it is notable that settlements in the western, central, and northern, regions of the eThekwini Metro including Berea, Bluff, Bayhead, Phoenix and Tshelimnyama had a high prevalence of COVID-19 during the 3rd wave as shown by the red and dark red colour on the map. It is also noticeable that areas in the southern parts of the Metro, such as Adams Mission and Umbumbulu, had relatively lower numbers of confirmed COVID-19 cases as shown by the peach colour.

The spatial pattern of coronavirus incidence is further shown in the spatial autocorrelation results **(Figure 3**) which show that settlement level COVID-19 incidence throughout the third wave in the eThekwini Metro exhibited a clustered spatial pattern. Notably, low COVID-19 cases were clustered in the southern and north-western parts of eThekwini while hotspots or high COVID-19 prevalence were more spatially concentrated in the central-eastern regions. However, in the northern and central regions of eThekwini- made up of a blend of industrial, informal and suburban settlements, such as Mgangeni SP, parts of Verulam and La Lucia- COVID-19 cases were randomly distributed, possibly due to mobility, temporal and spatial lags in reporting and variations in testing practices (Gibson and Rush, 2020).

Noticeably, the lack of COVID-19 testing in the northern and central informal settlements of eThekwini could have distorted the true distribution of COVID-19 cases and thus, resulted in the apparent geographic randomness of COVID-19 incidence in these areas (Shariati *et al*., 2020). Additionally, during the research period, which was the 3rd wave between May 02 to September 11 2021, lockdown levels 3 and 4 were in place (SAcoronavirus, 2021). Consequently, it is possible that the implementation of strict COVID-19 regulations could have reduced COVID-19 spread in the above-mentioned areas, thus, resulting in a statistically insignificant spatial pattern of COVID-19 incidence (Arashi *et al*., 2020). Although this study utilised data spanning a shorter period, the results relating to the random spread of COVID-19 cases in some areas, were similar to those of other studies, such as that of Arashi *et al*. (2020) who modelled and predicted COVID-19 incidence for South Africa and noted a significant decline (randomness) in COVID-19 incidence when strict lockdown regulations were introduced.

Overall, the results of Moran's *I* (spatial autocorrelation) clearly illustrate that settlement level COVID-19 incidence during the 3rd wave in eThekwini was spatially autocorrelated; and varied based on the geographic point of observation. The results mentioned above provide key insights into the strategies that are effective for managing COVID-19 and other communicable diseases, including the implementation of lockdowns to contain pandemics (Arashi *et al*., 2020; Mollalo *et al*., 2020, Mollalo *et al*., 2021). Further analysis using OLS, SEM, SLM, GWR and MGWR models exposed some of the significant predictors that steer COVID-19 incidence spatial variation and the effects thereof.

5.2. Global and local models for COVID-19 incidence modelling

The OLS model showed that there was no data multicollinearity between the 10 potential determinants of COVID-19 incidence as all VIF values were comparatively small, and as such, the model results accurately show the relation between the determinants and the dependent variable, which is the occurrence of COVID-19, is free of bias. Additionally, according to the OLS model, it is noticeable that positive linear associations between the predictors and COVID-19 prevalence were dominant as 6 of the coefficients were greater than zero (**Table 2**). Notably, out of the ten independent variables, only three factors were statistically significant as their p-values were lower than 0.05, and included population density, economic wellness score and

the number of TB cases. It is noticeable that various other studies collected relatively more variables but ended up with a much smaller number of modelled variables, suggesting that the potential determinants of COVID-19 incidence are not exhaustive (Mollalo *et al*., 2020). For instance, Mollalo *et al*. (2020) collected 35 variables and only modelled 4 variables while Dutta *et al*. (2021) collected 19 variables and only selected 3 for the models.

For this study, all the other variables other than population density, economic wellness score and the number of TB cases were statistically insignificant as their p-values were greater than 0.05. Consequently, as per 3rd wave confirmed cases data for eThekwini, gender, age, vaccination coverage, health facility density, pollution, poverty, and the vulnerability index were not determinants of COVID-19 incidence and did not significantly impact spatial variations in COVID-19 prevalence. The above is consistent on all the models as seen by the p-values and matches with outcomes from other studies, such as that of Dutta *et al*. (2021) who only found significant spatial relationships between indicators of population density and economic wellness on COVID-19 incidence in India (**Table 3**, **Table 4** and **Table 5**). However, other studies, including that of Ma *et al*. (2020) and Hassaan *et al*. (2021) found significant positive spatial associations between pollution, vaccination, and poverty on COVID-19 prevalence and mortality in Wuhan and African countries, respectively. Consequently, it is possible that the implemented lockdown regulations during the study period significantly reduced COVID-19 incidence at the settlement level within the eThekwini Metro resulting in pollution, poverty and demographic factors (age and gender) having no significant impact on the spatial spread of COVID-19 occurrences in the eThekwini Metro (Arashi *et al*., 2020; Mollalo *et al*., 2020).

The above also applies to vaccination coverage, which exhibited no significant spatial correlation with COVID-19 incidence (Arashi *et al*., 2020). Nevertheless, there is still a need to explore the effect of other pollution, poverty, age, and gender indicators on COVID-19 incidence within the eThekwini settlements as other studies suggest that the aforementioned factors have a substantial effect on COVID-19 incidence (Ma *et al*., 2020; Ramírez-Aldana *et al*., 2020). To illustrate, Ramírez-Aldana *et al*. (2020) found the number of people over 60 years and the average temperature to be significant predictors of COVID-19 incidence in Iran. Subsequently, exploring other indicators of pollution, age, gender, and poverty could potentially alter the results

obtained in this study (Ramírez-Aldana *et al*., 2020). Nevertheless, the obtained results still provide useful insights into the key predictors of COVID-19 occurrence at granular levels within the eThekwini Metro and illustrate the importance of monitoring population density, economic wellness score and the number of TB cases to reduce the spread of COVID-19 in the eThekwini settlements.

Moreover, it is noticeable that the study found health facility density to have an insignificant impact on the prevalence of COVID-19 within the eThekwini settlements. This could be attributed to healthcare capacity (the availability of healthcare personnel and resources) being a more substantial predictor of COVID-19 incidence compared to the presence or density of health facilities (Mollalo *et al*., 2020; Ramírez-Aldana *et al*., 2020). The above is in line with findings from Mollalo *et al*. (2020) and Ramírez-Aldana *et al*. (2020) who noted that the number of healthcare workers had a positive correlation with the occurrence of COVID-19 due to a larger number of healthcare workers being deployed in areas with higher population density, and consequently higher COVID-19 prevalence (Rahman *et al*., 2021). Subsequently, for further analysis of COVID-19 incidence within the eThekwini Metro settlements, the impact of the number of healthcare workers on COVID-19 incidence could be explored. Additionally, the insignificant impact of the COVID-19 vulnerability index on COVID-19 incidence noted within the eThekwini Metro suggests a need for further analysis of the effect of COVID-19 vulnerability over a longer temporal scale to inform COVID-19 management strategies (Mollalo *et al*., 2020; Dutta *et al*., 2021).

In addition to the above, while population density, economic wellness score, and TB were found to be significant predictors of COVID-19 in eThekwini and no multicollinearity existed, these variables are still interactive and understanding their interactive effects can provide more nuanced understanding of the dynamics of COVID-19 incidence (Rahman et al., 2021). To illustrate, areas with high population density are often associated with overcrowding, which can exacerbate the incidence and spread of COVID-19 (Irandoost *et al.,* 2023). However, areas with higher wellness scores are more likely to have access to improved healthcare services, and living conditions, along with more resources to put in place preventative measures for COVID-19 spread. This could potentially reduce the incidence of COVID-19 regardless of the high population density (Dutta *et al.,* 2021; Irandoost *et al.,* 2023). This was also the case in the study of Siddique *et al.* 2023, who found that COVID-19 incidence was

comparatively lower in areas with higher income average compared to those that were poverty stricken, despite both areas having comparable population density in the USA (Siddique *et al.,* 2023).

Additionally, the risk of faster transmission and incidence of COVID-19 due to compromised immune systems is relatively higher in areas with high cases or rates of comorbidities, such as TB (Mahara et al., 2018; Dutta *et al.,* 2021). However, higher wellness scores have the potential to counter the effect of comorbidities on COVID-19 incidence as they are often associated with improved access to healthcare services and living conditions, and resources for prevention measures, including better hygiene practices, which could potentially reduce COVID-19 incidence despite high levels TB. Notably, the above-mentioned interaction effects are explored further in the sections below in the context of the study area.

Additionally, it is notable that the OLS model explains 39.8% of the variation in the data, which alludes to the presence of a moderate effect on COVID-19 incidence by the significant predictors. Comparatively, the SEM and SLM models, which are global models like the OLS model, exhibited improved performance as they explained 39.9% and 39.8% of the variation in the occurrence of COVID-19 and adequately accounted for the potential predictors of COVID-19 incidence as seen from the smaller AIC values (**Table 3**). Consequently, models that account for spatial dependence, such as the SEM and SLM models, provide more valuable results when analysing and modelling COVID-19 incidence at micro-levels (Dutta *et al*., 2021). The above is corroborated by Dutta *et al*. (2021) and Mollalo *et al*. (2020) who noted that the SEM and SLM models outperformed the OLS model when analysing COVID-19 incidence in India and the USA, respectively.

Additionally, it is noticeable that the local level models, GWR and MGWR, explained 71.6% and 72.5% of the COVID-19 incidence variance, respectively (**Table 4** and **Table 5**). Moreover, the GWR and MGWR models had the lowest AIC values in comparison to the other models (**Table 4** and **Table 5**). Additionally, the MGWR model, which analyses spatial autocorrelation at multiple spatial scales, exhibited the best overall fit for the data and explanatory power compared to the other models as it had the lowest AIC value and the largest local R^2 value (**Table 5**). Consequently, local level models that account for spatial dependence provide better modelling of the

association between potential predictors and COVID-19 incidence at micro-levels as noted in various other studies (Saffary *et al*., 2020; Hassaan *et al*., 2021; Kindi *et al*., 2021; Dutta *et al*., 2021).

5.3. Population density and COVID-19 incidence spatial dependence

Moreover, spatial autocorrelation was utilised to analyse the spatial correlation of COVID-19 incidence on the three significant explanatory variables, which were economic wellness score, population density, and number of TB cases at micro-levels. Notably, population density is positively associated with COVID-19 incidence at micro-levels as shown by the clustered spatial autocorrelation (**Figure 4**). Similarly, economic wellness score and the number of TB cases have significant clustered patterns with COVID-19 at micro-levels as shown by the spatial autocorrelation results (**Figure 5** and **Figure 6**). Consequently, the above implies that in areas where the values of these determinants are high, there will be high COVID-19 prevalence. The above is in line with findings from other literature, including Dutta *et al*. (2021) who found a positive association between COVID-19 prevalence and population density and Mollalo *et al*. (2020) who found a positive association between COVID-19 incidence and income inequality.

When looking at population density, high clustering with COVID-19 incidence is notable in the central and north-eastern regions of the research site covering Durban Central, KwaMashu and Umlazi. The above-mentioned areas are composed of informal and township settlements with high populations (eThekwini Municipality, 2020). Thus, the strong positive relation between population density and the occurrence of COVID-19 can be ascribed to higher population density exacerbating the probability and frequency of COVID-19 spread through reducing social distancing (increasing contact) thus, heightening the incidence of COVID-19 (Dutta *et al*., 2021; Ahasan *et al*., 2022). Additionally, significant Low-Low clusters are dominant in the western and southern areas of the Metro covering Pinetown and Umbumbulu. The above results are similar to those obtained in several other studies including that of Dutta *et al*. (2020) and Mollalo *et al*. (2020) who noted positive associations between COVID-19 incidence and population density in India and the USA, correspondingly. However, in other studies, such as the study done by Irandoost *et al*. (2023), there was a lack of spatial association between population density and the occurrence of COVID-19 cases (Irandoost *et al*., 2023). Notably, Irandoost *et al*. (2023) utilised the

Pearson correlation coefficient and illustrated a lack of a strong spatial correlation between COVID-19 prevalence and population density per hectare and a strong correlation between COVID-19 incidence per 1000 people and population density. Consequently, it is also possible that a more granular level analysis, such as at a neighbourhood or street level, would have produced more robust results regarding the geographic relationship between COVID-19 incidence and population density (Irandoost *et al*., 2023). Nevertheless, the obtained results correspond with existing literature and provide valuable insights into the key predictors of COVID-19 incidence at a settlement level in the eThekwini Metro.

5.4. Economic wellness score and COVID-19 incidence spatial dependence

For the spatial autocorrelation of economic wellness score and COVID-19 incidence, Low-Low clusters are more dominant in the southern and north-western parts of the eThekwini Metro in rural settlements such as Umbumbulu, Adams Mission, Inkangala Mgangeni SP, and KwaNgcolosi while high COVID-19 incidence and economic wellness score are notable on the central and eastern parts of the Metro covering Pinetown, Newlands West, Westville, and Umhlanga; which are more urban settlements with more affluent economics and higher populations (**Figure 5**). Concerning the economic wellness score, it is possible that areas that are more affluent exhibit more movements and inflows of various populations which has the potential to increase the risk of COVID-19 infections and incidence (Hassaan *et al*., 2021; Rahman *et al*., 2021). Additionally, in areas with lower economic wellness scores, such as Umbumbulu and Mgangeni SP, there was potentially a lack of access to COVID-19 testing facilities, triggering reduced reporting of COVID-19 cases and thus, lower COVID-19 incidence (Hassaan *et al*., 2021). Notably, similar findings were presented in other studies, such as that of Rahman *et al*. (2021) who noted a positive spatial correlation between COVID-19 incidence and monthly consumption- an economic wellness indicator- in Bangladesh, with lower COVID-19 incidence being dominant in the north-western part of Bangladesh, which had a lower economic wellness. Subsequently, from the results presented regarding economic wellness and COVID-19 incidence within the eThekwini settlements, it is clear that economic wellness is a key determinant of COVID-19 prevalence, which should be accounted for when developing mitigation strategies for COVID-19 and other communicable diseases (Ahasan *et al*., 2022).

5.5. Number of TB cases and COVID-19 incidence spatial dependence

Noticeably, the spatial association between the number of TB cases and COVID-19 incidence is only significant around Durban Central, Berea and some parts of Umlazi where High-High and Low-High clusters are notable, and in the southern parts of the study area in Madundube SP where High-Low clusters (**Figure 6**). The spatial patterns of TB cases and COVID-19 incidence, particularly the clustering of high values, could be attributed to the presence of industrial areas which have positive associations with TB due to high population densities, poor ventilation and exposure to indoor pollutants (Kootbodien *et al*., 2018).

Another potential explanation for the noted pattern is the lack of reporting of TB incidence particularly during the study period due to lockdown regulations, which would have resulted in an incomplete reporting of TB incidence and thus, the apparent weak association between COVID-19 incidence and the number of TB cases (**Figure 6**) (Kootbodien *et al*., 2018). Notably, the study undertaken by Berra *et al*. (2022) in Brazil highlighted a positive spatial association between the occurrence of COVID-19 and the occurrence of TB cases. Although it is clear that the number of TB cases was generally positively spatially associated with COVID-19 incidence at a settlement level during the 3rd wave in eThekwini, there exists a need for data with a higher temporal and spatial resolution to fully explore the impact of the number of TB cases on COVID-19 incidence, and infectious diseases as a whole (Kootbodien *et al*., 2018).

5.6. Effects of population density on COVID-19 incidence

The effects of the significant variables on COVID-19 incidence were also analysed using GWR and MGWR models. It was noted that the spatial autocorrelation between micro-level COVID-19 incidence and population density in the central, northern, south-western, and western regions of the eThekwini Metro in settlements such as Tongaat, Ottawa, Umbumbulu, parts of Umlazi and Pinetown was very strong for both models, suggesting that high population densities tend to be associated with a higher number of COVID-19 cases (**Figure 7**). In some southern central and north-western settlements such as Madundube SP, Umkomaas, and Mgangeni lower GWR and MGWR coefficients were noticeable, alluding to strong negative associations between the density of populations and COVID-19 incidence in these areas. Noticeably, the effect of population density on COVID-19 incidence did not exhibit a well-defined pattern, potentially due to the impacts of human movement and the relatively broad

geographic scale in comparison to other studies, such as that of Irandoost *et al*. (2023), which considered COVID-19 incidence per 1000 people (Saffary *et al*., 2020). Nevertheless, the GWR and MGWR results demonstrated the varying effects of population density on COVID-19 incidence at a settlement level within the eThekwini Metro, thus providing valuable information for targeted COVID-19 interventions and contributing to the literature on spatial patterns of infectious diseases.

5.7. Effects of economic wellness score on COVID-19 incidence

Comparatively, the spatial effects of the economic wellness score in describing COVID-19 incidence showed a well-defined spatial pattern with higher coefficients being on the coastal edge and slowly decreasing moving inland, especially on the MGWR model (**Figure 8**). Similarly, Rahman *et al*. (2021) noted a positive spatial relationship and uniform distribution pattern between monthly consumption (an economic wellness indicator) and COVID-19 incidence in a study conducted in Bangladesh. Consequently, a high economic wellness score is linked with higher COVID-19 incidence along the south-eastern and eastern areas of the coast extending into the upper central settlements including Durban Central, Berea, Bluff, Verulam, Umlazi and Westville. Thus, as per the 3rd wave confirmed COVID-19 cases data, economic wellness score is positively correlated with COVID-19 incidence at micro-levels on both the GWR and MGWR models. This implies that in addition to population density and the sum of TB cases (in a few areas noted in the sections above), high COVID-19 incidence along the south-eastern and eastern regions of the eThekwini coast was also due to a high economic wellness score.

5.8. Effects of TB prevalence on COVID-19 incidence

The results for the spatial dependence of micro-level COVID-19 incidence on TB prevalence (number of TB cases) illustrate that there is a strong positive correlation between the sum of TB cases and COVID-19 incidence, with the strongest associations being notable in the southern and eastern areas of the eThekwini Metro, in areas such as Madundube SP, Berea and Clairwood, while smaller positive associations are notable on the western and northern parts of the Metro in areas such as Tongaat and Mgangeni (**Figure 9**). Additionally, the slightly inconsistent pattern of the association between TB prevalence and COVID-19 incidence could be attributed to the limited availability of data on TB cases at granular spatial scales, thus resulting

in somewhat sporadic patterns. Notably, patterns similar to the above were also highlighted in other studies, such as that of Dutta *et al*. (2021), who noticed an inconsistent pattern in the effects of urbanisation in explaining COVID-19 incidence using the GWR model, possibly due to data limitations (Dutta *et al*., 2021). Comparatively, Berra *et al*. (2022) noted a more consistent pattern of positive spatial correlation between COVID-19 incidence and TB prevalence due to the utilisation of data with a higher temporal resolution. The results demonstrate that TB prevalence was a driver of COVID-19 spread in some eastern and southern settlements of the Metro, such as Berea, in addition to population density and economic wellness. Additionally, the results provide valuable insights into the effects of the sum of TB cases on micro-level COVID-19 incidence in the eThekwini Metro, notwithstanding limitations of data availability and temporal resolution.

5.9. GWR and MGWR Local R^2

In addition to the above, from the maps of local R^2 for the MGWR and GWR models, it is notable that the eastern and central portions of eThekwini Metro in settlements such as Berea, Bluff, Westville, Newlands West and parts of Umlazi exhibited the highest association and explanatory power between COVID-19 incidence and the predictors of the occurrence of COVID-19 which were statistically significant (**Figure 10**). Noticeably, the GWR and MGWR local R^2 values were similar to those of other studies, such as that of Mollalo *et al*. (2020) which had GWR and MGWR local R^2 values ranging from 0 to 0.75. Subsequently, the strongest relationships between COVID-19 incidence and its explanatory variables are situated in the areas mentioned above, which could be underlined for intervention strategies, such as deciding on the strategies for enforcing mask-wearing and social distancing regulations and could provide valuable insights into COVID-19 the impacts thereof at granular scales. Additionally, lower correlations were s on the western, northern, and southern edges of the Metro in settlements such as Tongaat, Mpumalanga, Umbumbulu and Mgangeni where the GWR and MGWR models exhibited lower explanatory power and weaker model fit. The above could be attributed to limited data availability, especially on the number of TB cases at smaller spatial scales, resulting in increased variability and reduced statistical power of the models (Brunsdon *et al*., 1996; Oshan *et al*., 2019; Berra *et al*., 2022).

Finally, from all the results presented above, it can then be said that during the third wave, COVID-19 hotspots were in areas such as Bayhead, Berea, and Bluff as a result of high population density, high economic wellness score and a high number of TB cases. However, it is notable that the association between the number of TB cases and COVID-19 incidence at the settlement level was comparatively low. This could be ascribed to the limited availability of granular level data on the number of TB cases. Additionally, it is possible that other indicators of comorbidities could have exhibited better explanatory power for COVID-19 incidence compared to TB prevalence (Hassaan *et al*., 2021). This is illustrated in the study undertaken by Saffary *et al*. (2020) which highlighted positive associations between COVID-19 incidence and comorbidities, including diabetes and obesity. In this context, it is clear that the settlement level relationship between the number of TB cases and COVID-19 incidence still requires further modelling and investigation along with the exploration of the relationship between other comorbidities, such as diabetes, on micro-level COVID-19 incidence. Additionally, modelling COVID-19 incidence and population density at a more granular level, such as at a street or neighbourhood level, could further strengthen insights on the spatial aetiology of COVID-19, and other similar epidemics along with exploring other health security indicators of COVID-19 incidence, including the number of healthcare workers.

5.10. Limitations of the study

This study was limited by the number of predictors collected, as only ten variables were utilised. As the potential predictors of COVID-19 incidence are not exhaustive, collecting more indicators for each potential predictor could have potentially provided more insights into the determinants of COVID-19 incidence at micro-levels. However, the lack of granular-level mobility, behavioural, and socioeconomic data collected through detailed granular-level primary data collection, which was not feasible in the study, limited a broader understanding of the explanatory variables of COVID-19 incidence and spread. Nevertheless, the utilised variables offer valuable insights into the relationship between socioeconomic factors and COVID-19 spread.

Additionally, the limited availability of data on the number of TB cases at settlement levels, and updated population estimates along with incorrect addresses on the COVID-19 cases data, reduced data density and subsequently, the statistical power of the models. Additionally, the data density and improved modelling power of the

study were further reduced by the temporal resolution of COVID-19 case data, as only one wave of data was utilised. Notably, although additional temporal data could provide an improved understanding of the pandemic dynamics across all waves, data for waves 1, 2 and 4 were inaccessible as most of it was not collected in a digital format or a format suitable for research analysis.

Additionally, the use of local spatial regression techniques, which have the limitation of not being able to handle large and complex datasets could have reduced the modelling power. Nevertheless, the models were still sufficient in providing a broad understanding of the determinants of COVID-19 incidence within the Metro. In addition to the above, model misspecification could have occurred due to the presence of interaction effects, introducing complexities that the GWR and MGWR models mat noy have been able to handle (Mollalo *et al.,* 2020). However, careful selection of the explanatory variables was undertaken to mitigate this issue.

Additionally, the local spatial models, GWR and MGWR, are sensitive to the choice of input parameters used (Dutta *et al.,* 2021). Therefore, minor changes in the input parameters could have introduced uncertainty in the interpretation of the relationships between COVID-19 incidence and the explanatory variables. Also, the generalisability of the results could have been compromised due to the susceptibility of the models to overfitting, which could be attributed to the use of small bandwidths (Zhang and Song, 2024). Therefore, adjustments to the models would be required to ensure the models perform optimally when new data is introduced. Nevertheless, the obtained results provide valuable insights into the spread of COVID-19 at granular levels within the eThekwini Metro as the models were simplified to include only a few explanatory variables, and model validation was undertaken to reduce overfitting and sensitivity. Moreover, more granular spatial scales, such as street or neighbourhood scales, were not covered in the study. Consequently, spatial dynamics of COVID-19 incidence and determinants, such as population density, at finer scales were not explored. Nevertheless, various approaches, such as regression and the exclusion of values were utilised to overcome the above-mentioned data limitations.

CHAPTER 6

Conclusion and Recommendations

6.1. Conclusion

This study aimed to explore the use of advanced geospatial analysis and modelling techniques in modelling COVID-19 incidence at micro-levels using the eThekwini Metropolitan area as a case study. One of the objectives of the study was to map the granular level distribution of COVID-19 over the 3rd wave in eThekwini Metro through Moran's *I* (spatial autocorrelation). This objective was achieved using Moran's *I* to map the spatial pattern of COVID-19 incidence at a granular level. Moran's *I* was computed using the data on the eThekwini settlement level number of confirmed cases. Both spatially and statistically insignificant values along with outliers were removed. Subsequently, spatial autocorrelation was utilised to demonstrate the spatial autocorrelation of COVID-19 prevalence, which had a Moran's *I* value of 0.14 and exhibited a clustered pattern, indicating the settlement level spatial autocorrelation of COVID-19 incidence. It was also illustrated that COVID-19 incidence exhibited positive associations and a clustered pattern within settlements in the Metro such as Bayhead and Bluff.

In addition to the above, comparing the applicability of global and local models in analysing micro-level COVID-19 incidence was another objective of the study. This objective was achieved using OLS, SEM, SLM, GWR and MGWR models for modelling ten potential explanatory variables with COVID-19 incidence. The models showed that population density, the prevalence of TB cases and economic wellness score were statistically significant predictors of settlement level COVID-19 incidence as they had p-values less than 0.05. The results illustrated that local level models that account for spatial dependence were more effective in identifying and measuring the main determinants of COVID-19 incidence at micro-level spatial scales, as the MGWR model (a local level multiscale model) showed the best performance and explained 72.5% of the variance in COVID-19 incidence, which was 33% more variance than the global models. Additionally, the low AIC value of the MGWR model (193.392) further demonstrated that local level models provide better explanatory power and overall fit to the data compared to global models when modelling COVID-19 incidence.

The third objective of the study was to analyse the spatial dependence of COVID-19 incidence on the explanatory variables using spatial autocorrelation. The results of the spatial autocorrelation illustrated that of the ten chosen explanatory variables, there were only three significant explanatory variables, which were population density, economic wellness score and number of TB cases. The three significant determinants were all positively spatially correlated with COVID-19 incidence and exhibited significant High-High clusters in areas such as Berea, Bluff, Umhlanga, and Umlazi.

Finally, the study also aimed to spatially model the effects of significant local-level determinants on COVID-19 incidence. The above was achieved through mapping the GWR and MGWR coefficients for population density, economic wellness score, and number of TB cases. The results indicated that population density had very strong positive associations with COVID-19 incidence in areas such as Umlazi and Bluff, and much weaker associations in the southern-central and north-western parts of the study area in settlements such as Madundube and Mgangeni. Additionally, the economic wellness score showed a well-defined spatial pattern with a stronger positive correlation with COVID-19 incidence in the coastal settlements. The correlation gradually lessens when moving towards the inland settlements. For TB prevalence, a slightly inconsistent spatial pattern (effect on COVID-19 incidence) was noted, due to the limited data availability. Higher coefficients were more notable on the southern and eastern regions of the study area while lower coefficients were dominant on the northern and western parts of the study area.

Conclusively, the study demonstrated the applicability of advanced geospatial models in analysing and modelling micro-level COVID-19 incidence and provided valuable insights into the status and predictors of COVID-19 incidence within the eThekwini Metro settlements. This information provided a basis for using advanced models to enhance spatial understanding and visualisation of COVID-19 distribution at granular levels, decrease bias of visual perceptions, and inform effective COVID-19 response strategies.

6.2. Recommendations

It is recommended that additional research using more potential explanatory variables of COVID-19 be utilised to further explore the capabilities of advanced geostatistical and geospatial models in COVID-19 studies. However, other geospatial models, such as Bayesian hierarchical models, should be added to explore the above, as they are able to handle large and complex datasets. Following from this, potential research questions could include modelling and analysis of COVID-19 spread at microlevels using Bayesian hierarchical models. Moreover, the generalisability of this study's findings could be enhanced through strengthening the modelling capabilities of the local spatial models by employing cross-validation techniques, sensitivity analysis, and simulation studies to assess the robustness of the models under varying conditions and improve their performance. Potential methodologies in this context include altering the spatial weights matrix for the SEM and SLM models to evaluate the impact of spatial dependence assumptions on the results, and the addition of perturbation simulations to the local spatial models to assess their ability to handle uncertainty and variability.

Additionally, increasing the temporal coverage and spatial resolution of the COVID-19 data could be beneficial for gaining more accurate and in-depth insights into the temporal and spatial dependency of the COVID-19 incidence at finer micro-level scales, such as at a street level. Consequently, the spatio-temporal analysis and modelling of COVID-19 incidence at microlevels within the eThekwini Metro could be a potential research question in this regard. Moreover, predictive modelling of COVID-19 incidence at granular levels, and other diseases, using local level spatial models should be explored to fill existing gaps in spatial epidemiology. Notably, in the context of this study, a potential approach of achieving the above would be to add cross-validation techniques, such as K-Fold cross-validation, to understand the models' predictive power and robustness.

References

Ahasan, R., Alam, M.S., Chakraborty, T. and Hossain, M.M. 2022. Applications of GIS and geospatial analyses in COVID-19 research: A systematic review. *F1000Research*, 11, 1–17.

Akinwumiju, A.S., Oluwafemi, O., Mohammed, Y.D. and Mobolaji, J.W. 2022. Geospatial evaluation of COVID-19 mortality: Influence of socio-economic status and underlying health conditions in contiguous USA. *Applied Geography*, 141, 102671, doi:10.1016/j.apgeog.2022.102671.

Arashi, M., Bekker, A., Salehi, M., Millard, S., Erasmus, B., Cronje, T. and Golpaygani, M. 2020. Spatial analysis and prediction of COVID-19 spread in South Africa after lockdown. *arXiv*, 2005, 09696, doi:10.48550/arXiv.2005.09596.

Barrett, F.A. 2000. August Hirsch: as critic of, and contributor to, geographical medicine and medical geography. *Medical History*, 44, 98–117.

Berra, T.Z., Ramos, A.C.V., Alves, Y.M., Tavares, R.B.V., Tartaro, A.F., do Nascimento, M.C., Moura, H.S.D., Delpino, F.M., de Almeida Soares, D., Silva, R.V. dos S., Gomes, D., Monroe, A.A. and Arcêncio, R.A. 2022. Impact of COVID-19 on Tuberculosis Indicators in Brazil: A Time Series and Spatial Analysis Study. *Tropical Medicine and Infectious Disease*, 7, 247–266.

Broadbent, A., Combrink, H. and Smart, B. 2020. COVID-19 in South Africa. *Global Epidemiology*, 2, 1–4.

Brunsdon, C., Fotheringham, A.S. and Charlton, M.E. 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28, 281–298.

Castro, R.R., Santos, R.S.C., Sousa, G.J.B., Pinheiro, Y.T., Martins, R.R.I.M., Pereira, M.L.D. and Silva, R. a. R. 2021. Spatial dynamics of the COVID-19 pandemic in Brazil. *Epidemiology & Infection*, 149, 1–9.

Chandra, S. and Sharma, M. 2024. Determinants of COVID-19 Prevalence Rate in Asia: A study using Spatial Analysis. *Indian Journal of Public Health Research & Development*, 15, 37506, doi:10.37506/q9ffyr61.

Clarke, K.C., McLafferty, S.L. and Tempalski, B.J. 1996. On epidemiology and geographic information systems: a review and discussion of future directions. *Emerging Infectious Diseases*, 2, 85–92.

Connor, H. 2022. John Graunt F.R.S. (1620-74): The founding father of human demography, epidemiology, and vital statistics. *Journal of Medical Biography*, 32, 57–69.

Cromley, E.K. 2019. Using GIS to address epidemiologic research questions. *Current Epidemiology Reports*, 6, 162–173.

Dutta, I., Basu, T. and Das, A. 2021. Spatial analysis of COVID-19 incidence and its determinants using spatial modeling: A study on India. *Environmental Challenges*, 4, 1–10.

Editorial. 2018. Epidemiology is a science of high importance. *Nature Communications*, 9, 1703–1704.

Esri, 2022a. *How OLS regression works*. Available from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-ols-regression-works.htm, accessed 23 August 2022.

Esri, 2022b. *Spatial Autocorrelation (Global Moran's I) (Spatial Statistics)*. Available from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/spatial-autocorrelation.htm, accessed 14 September 2022.

eThekwini Municipality, 2020. *eThekwini Metropolitan KZN*. Available from https://www.cogta.gov.za/ddm/wp-content/uploads/2020/07/Metro-Profile_Ethekwini, accessed 29 May 2022.

Fatima, M., O'Keefe, K.J., Wei, W., Arshad, S. and Gruebner, O. 2021. Geospatial analysis of COVID-19: A scoping review. I*nternational Journal of Environmental Research and Public Health*, 18, 1–14.

Franch-Pardo, I., Napoletano, B.M., Rosete-Verges, F. and Billa, L. 2020. Spatial analysis and GIS in the study of COVID-19. A review. *Science of The Total Environment*, 739, 1–10.

Gibson, L. and Rush, D. 2020. Novel coronavirus in Cape Town informal settlements: Feasibility of using informal dwelling outlines to identify high risk areas for COVID-19 transmission from a social distancing perspective. *JMIR Public Health and Surveillance*, 6, 1–9.

He, Y., Seminara, P.J., Huang, X., Yang, D., Fang, F. and Song, C. 2023. Geospatial modeling of health, socioeconomic, demographic, and environmental factors with covid-19 incidence rate in arkansas, us. *ISPRS International Journal of Geo-Information*, 12, 45–77.

Hassaan, M.A., Abdelwahab, R.G., Elbarky, T.A. and Ghazy, R.M. 2021. GIS-Based analysis framework to identify the determinants of COVID-19 incidence and fatality in Africa. *Journal of Primary Care & Community Health*, 12, 1–12.

Irandoost, K., Alizadeh, H., Yousefi, Z. and Shahmoradi, B. 2023. Spatial analysis of population density and its effects during the COVID-19 pandemic in Sanandaj, Iran. *Journal of Asian Architecture and Building Engineering*, 22, 635–642.

Jagarnath, M., Thambiran, T. and Gebreslasie, M. 2020. Heat stress risk and vulnerability under climate change in Durban Metropolitan, South Africa—identifying urban planning priorities for adaptation. *Climatic Change*, 163, 807–829.

Kargon, R. 1963. John Graunt, Francis Bacon, and the Royal Society: The Reception of Statistics. *Journal of the History of Medicine and Allied Sciences*, 18, 337–348.

Kindi, K.M.A., Al-Mawali, A., Akharusi, A., Alshukaili, D., Alnasiri, N., Al-Awadhi, T., Charabi, Y. and Kenawy, A.M.E. 2021. Demographic and socioeconomic determinants of COVID-19 across Oman - A geospatial modelling approach. *Geospatial Health*, 16, 145–160.

Kootbodien, T., Wilson, K., Tlotleng, N., Ntlebi, V., Made, F., Rees, D. and Naicker, N. 2018. Tuberculosis Mortality by Occupation in South Africa, 2011–2015. *International Journal of Environmental Research and Public Health*, 15, 2756, doi:10.3390/ijerph15122756.

Krieger, N. 2003. Place, space, and health: GIS and epidemiology. *Epidemiology*, 14, 384–385.

KZN DoH, 2022. *Fast-track cities- eThekwini COVID-19 dashboard*. Available from https://www.fast-trackcities.org/cities/ethekwini-covid, accessed 27 April 2022.

Lin, C.-H. and Wen, T.-H. 2022. How Spatial Epidemiology Helps Understand Infectious Human Disease Transmission. *Tropical Medicine and Infectious Disease*, 7, 164, doi: 10.3390/tropicalmed7080164.

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J. and Luo, B. 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of The Total Environment*, 724, 138226, doi: 10.1016/j.scitotenv.2020.138226.

Mahara, G., Yang, K., Chen, S., Wang, W. and Guo, X. 2018. Socio-economic predictors and distribution of Tuberculosis incidence in Beijing, China: A study using a combination of spatial statistics and GIS technology. *Medical Sciences*, 6, 26, doi: 10.3390/medsci6020026.

Maharaj, A. and Reddy, P. 2020. Local government's economic response to COVID-19: The case of eThekwini City Council in Durban, South Africa. *Alternation*, 32, 197–223.

Manda, S.O.M., Darikwa, T., Nkwenika, T. and Bergquist, R. 2021. A spatial analysis of COVID-19 in African countries: Evaluating the effects of socio-economic vulnerabilities and neighbouring. *International Journal of Environmental Research and Public Health*, 18, 1–15.

Mbambo, S.B. and Agbola, S.B. 2020. The impact of the COVID-19 pandemic in townships and lessons for urban spatial restructuring in South Africa. *African Journal of Governance & Development*, 9, 329–351.

Mishra, A. and Kumar, J. 2021. Medical geographic information systems (Medical GIS): A review. *Agricultural Science: Research and Reviews*, 1, 71–78.

Mohi, S.M. 2011. GIS at public health. King Fahd University of petroleum and minerals, Dhahran.

Mollalo, A., Vahedi, B. and Rivera, K.M. 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, 728, 138884, doi: 10.1016/j.scitotenv.2020.138884.

Mollalo, A., Mohammadi, A., Mavaddati, S. and Kiani, B. 2021. Spatial Analysis of COVID-19 Vaccination: A Scoping Review. *International Journal of Environmental Research and Public Health*, 18, 12024, doi: 10.3390/ijerph182212024.

Morabia, A. 2015. Has Epidemiology Become Infatuated with Methods? A Historical Perspective on the Place of Methods During the Classical (1945–1965) Phase of Epidemiology. *Annual Review of Public Health*, 36, 69–88.

Mukandavire, Z., Nyabadza, F., Malunguza, N.J., Cuadros, D.F., Shiri, T. and Musuka, G. 2020. Quantifying early COVID-19 outbreak transmission in South Africa and exploring vaccine efficacy scenarios. *PLOS ONE*, 15, 1–11.

Oshan, T.M., Li, Z., Kang, W., Wolf, L.J. and Fotheringham, A.S. 2019. mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *ISPRS International Journal of Geo-Information*, 8, 269, doi: 10.3390/ijgi8060269.

Pillay, L., Amoah, I.D., Deepnarain, N., Pillay, K., Awolusi, O.O., Kumari, S. and Bux, F. 2021. Monitoring changes in COVID-19 infection using wastewater-based epidemiology: A South African perspective. *Science of The Total Environment*, 786, 147273, doi: 10.1016/j.scitotenv.2021.147273.

Rahman, Md.H., Zafri, N.M., Ashik, F.R., Waliullah, M. and Khan, A. 2021. Identification of risk factors contributing to COVID-19 incidence rates in Bangladesh: A GIS-based spatial modeling approach. Heliyon, 7, e06260, doi:10.1016/j.heliyon.2021.e06260.

Ramírez-Aldana, R., Gomez-Verjan, J.C. and Bello-Chavolla, O.Y. 2020. Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. *PLoS Neglected Tropical Diseases*, 14, e0008875, doi: 10.1371/journal.pntd.0008875.

Reddy, P.S. and Govender, N. 2019. Effectiveness of governance towards digitalisation at eThekwini Metropolitan Municipality in KwaZulu-Natal province, South Africa. *Africa's Public Service Delivery and Performance Review*, 7, 1–9.

Rytkönen, M.J.P. 2004. Not all maps are equal: GIS and spatial analysis in epidemiology. I*nternational Journal of Circumpolar Health*, 63, 9–24.

SAcoronavirus, 2021. *What Does South Africa's COVID Vaccine Roll-out Plan Say? - SA Corona Virus Online Portal*. Available from https://sacoronavirus.co.za/2021/01/12/what-does-south-africas-covid-vaccine-roll-out-plan-say, accessed 25 November 2023.

Saffary, T., Adegboye, O.A., Gayawan, E., Elfaki, F., Kuddus, M.A. and Saffary, R. 2020. Analysis of COVID-19 cases' spatial dependence in US counties reveals health inequalities. *Frontiers in Public Health*, 8, 579190, doi: 10.3389/fpubh.2020.579190.

Sangkham, S., Thongtip, S. and Vongruang, P. 2021. Influence of air pollution and meteorological factors on the spread of COVID-19 in the Bangkok Metropolitan Region and air quality during the outbreak. *Environmental Research*, 197, 111104, doi: 10.1016/j.envres.2021.111104.

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

Shrestha, S., Bauer, C.X.C., Hendricks, B. and Stopka, T.J. 2022. Spatial epidemiology: An empirical framework for syndemics research. *Social Science & Medicine*, 295, 113352, doi: 10.1016/j.socscimed.2020.113352.

Siddique, A.B., Haynes, K.E., Kulkarni, R. and Li, M.H. 2023. Regional poverty and infection disease: early exploratory evidence from the COVID-19 pandemic. *The Annals of Regional Science*, 70, 209–236.

Snow, J. 1856. Cholera and the Water Supply in the South Districts of London in 1854. *Journal of Public Health, and Sanitary Review*, 2, 239–257.

Statistics South Africa, 2020. *Protecting South Africa's elderly*. Available from https://www.statssa.gov.za/?p=13445, accessed 22 April 2022.

Statistics South Africa, 2022. South Africa: *Population*. Available from https://www.statssa.gov.za/?m=2022, accessed 28 May 2022.

Stevenson, L.G. 1965. Putting Disease on the Map: The Early Use of Spot Maps in the Study of Yellow Fever. *Journal of the History of Medicine and Allied Sciences*, 20, 226–261.

Susser, M. 1989. Epidemiology Today: 'A Thought-Tormented World.' *International Journal of Epidemiology*, 18, 481–488.

Thammaboribal, P., Tripathi, N.K., Junpha, J., Lipilert, S. and Wongpituk, K. 2024. Examining the Correlation between COVID-19 Prevalence and Patient Behaviors, Healthcare, and Socioeconomic Determinants: A Geospatial Analysis of ASEAN Countries. *International Journal of Geoinformatics*, 20, 95–112.

Tsoucalas, G., Laios, K., Karamanou, M. and Androutsos, G. 2013. The Thasian epidemic of mumps during the 5th century BC. *Infez Med*, 21, 149–50.

Waldhör, T. 1996. The Spatial Autocorrelation Coefficient Moran's I Under Heteroscedasticity. *Statistics in Medicine*, 15, 887–892.

Wang, Y., Chen, X. and Xue, F. 2024. A Review of Bayesian Spatiotemporal Models in Spatial Epidemiology. *ISPRS International Journal of Geo-Information*, 13, 97–126.

Zhang, Z., Li, Z. and Song, Y. 2024. On ignoring the heterogeneity in spatial autocorrelation: consequences and solutions. *International Journal of Geographical Information Science*, 2391981, doi: /10.1080/13658816.2024.2391981.

Appendices

Appendix 1: Economic wellness scores and descriptions

Score	*Description*
0	No data
1	Very poor; low-income class
2	Low emerging middle
3	Lower middle
4	Upper middle
5	Emerging (affluent)

Appendix 2: Primary dominant gender values

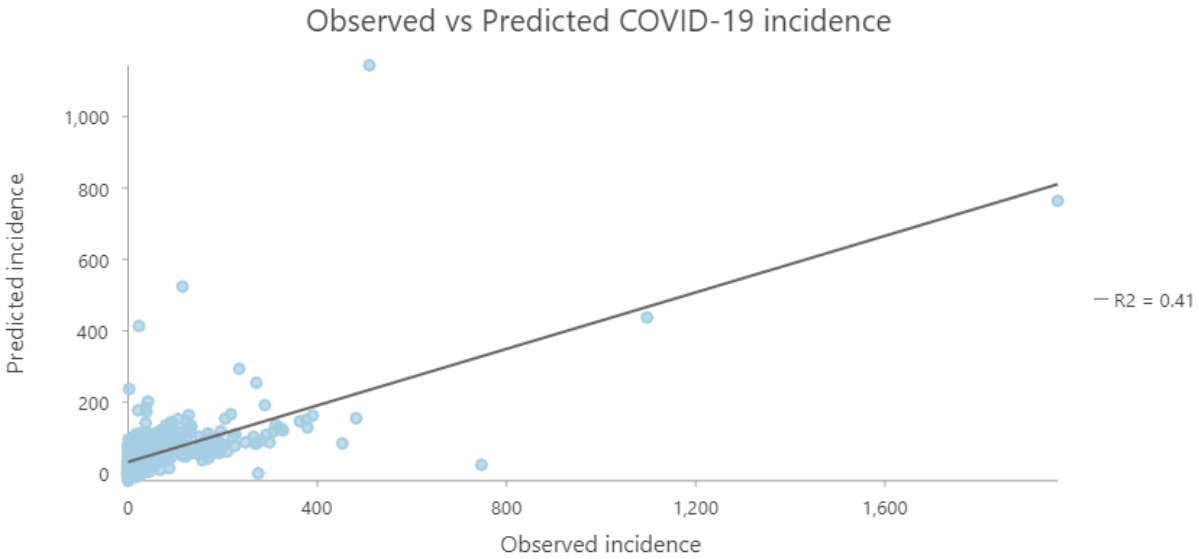
Gender	*Values*
Female	1
Male	0

For the primary dominant gender, male was assigned a value of 0 while female was given a value of 1.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-11.825590	14.590320	-0.810509	0.417948	14.113604	-0.837886	0.402407	--------
SUM_POP_KM2	0.000205	0.000023	8.841495	0.000000*	0.000130	1.579413	0.114768	1.538109
MEAN_WELNESS	24.745952	3.784852	6.538156	0.000000*	4.343676	5.697007	0.000000*	1.029584
EVDS_TOTAL	0.005104	0.000708	7.205199	0.000000*	0.003619	1.410385	0.158943	1.553152
DGENDER	-9.145344	7.643838	-1.196433	0.231988	6.604223	-1.384772	0.166635	1.006844

Appendix 3: Summary of OLS results



Appendix 4: Observed vs predicted COVID-19 incidence based on the number of confirmed COVID-19 cases. The GWR showed reasonably accurate predictions.