

https://dspace.org/

Spatio-temporal modelling of COVID-19 incidence in Bulawayo, Zimbabwe

Moyo, Velile

2022

https://hdl.handle.net/10539/34562

Downloaded from a DSpace repository. https://dspace.org/



SPATIO-TEMPORAL MODELLING OF COVID-19 INCIDENCE IN BULAWAYO, ZIMBABWE

Master's Research Report

By

Velile Moyo

1880716

Supervisor

Dr. Iqra Atif

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Geographic Information Systems and Remote Sensing in the School of Geography, Archaeology and Environmental Studies, Faculty of Science, University of the Witwatersrand

2021

Candidate's declaration

I declare that this research report is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



15th day of June 2022 at Bulawayo, Zimbabwe

ABSTRACT

The study mapped and evaluated the spatial and temporal patterns of COVID-19 in Bulawayo and identified the spatiotemporal distribution characteristics and changing trends of cases. From April 2020 to August 2021, the spatiotemporal distribution characteristics and shifting trends of cases were found. A time-series analysis of COVID-19 reported cases was done to determine trends, their direction, and size using the Mann Kendal trend test and the Sen slope. For the first three waves of the disease, a hotspot analysis, and a Cluster and Outlier analysis were used to identify the spatial distribution and clustering of COVID-19 cases. A Space-Time Cube of COVID-19 cases was then created to enable a 3D analysis of COVID-19 in space and time. An Emerging hotspot analysis and a Local Outlier Analysis were used to identify the hotspot clusters temporally and determine the trends of the hotspots. COVID-19 occurrences were then forecasted spatially for September 2021 using Exponential smoothing and Forest-based forecasting. An evaluation by location was done to produce a combination of the two prediction models. Hotspots were identified in the eastern parts of Bulawayo throughout the period and no clustering of the cold spot was observed using the 2D hotspot analysis. The 3D Local outlier analysis showed that fourteen (48,28%) of the locations had multiple patterns, nine (31,03%) locations contained low-low clusters, one (3,45%) location was a high-high cluster and five of the wards (17,24%) in Bulawayo were insignificant. The Emerging hotspot analysis identified two sporadic hotspots southeast of Bulawayo and thirteen consecutive hot spots concentrated mainly in the central parts of Bulawayo. Using the RMSE, the Forest-based forecasting was a better prediction model than the exponential smoothing forecast in predicting the spatiotemporal distribution of COVID-19 in Bulawayo. Thirteen out of twenty-nine wards in Bulawayo were predicted using the exponential smoothing forecast and sixteen of the wards used Forest-based forecasting. The findings of this study prove the significance of geospatial mapping which can quickly present and generate maps by highlighting the geographical and temporal variation of vulnerable areas. The study also proves the possibility to predict COVID-19 cases for future time steps using the exponential smoothing forecast and the Forest-based forecasting. The study recommends the use of geospatial techniques in monitoring and modeling the spatial trend and spread of pandemics that could help in targeted testing, prioritizing vaccination areas, and optimizing the allocation of limited resources.

Keywords: COVID-19; hotspot; Space-Time Cube; Emerging hotspot; forecasting

DEDICATION

A special feeling of gratitude to my loving mother, Doreen Moyo, who kept motivating me and offering words of encouragement.

My brothers Methembe and Qhelani have never left my side. I also dedicate this work to Connie and Elphas Dzimiri who have supported me throughout the process.

ACKNOWLEDGEMENTS

I would like to give special thanks to my supervisor Dr. Iqra Atif, who offered me continuous support throughout my master's journey. Her encouragement, patience, and knowledge have made this work possible.

I would like to thank the School of Geography, Archaeology, and Environmental Studies at the University of the Witwatersrand for providing me with the perfect academic environment to pursue my dreams.

Finally, I would like to thank the city council of Bulawayo for allowing me to conduct the research and use the data they had compiled.

TABLE OF CONTENTS

LIST OF ACRONYMS…….....i

LIST OF FIGURES ….....ii

LIST OF TABLES ….....iii

1. INTRODUCTION 1
1.1 BACKGROUND 1
1.2 RESEARCH PROBLEM STATEMENT 1
1.3 AIM AND OBJECTIVES.....2
1.3.1 Aim 2
1.3.2 Objectives 2
1.4 RESEARCH QUESTIONS 2
1.5 RESEARCH RATIONALE.....2
2. LITERATURE REVIEW 3
3. MATERIALS AND METHODS.....10
3.1 Study area.....10
3.2 Research design 11
3.3 Data requirement.....12
3.4 Methodology 12
3.4.1 Mann- Kendall statistic 13
3.4.2 Sen slope estimator 13
3.4.3 Space-Time cube.....14
3.4.4 Anselin Local Moran's I 15
3.4.5 Getis-Ord Gi*.....15
3.4.6 Holt-Winters (Triple exponential) smoothing 16
3.4.7 Random forests 17
3.4.8 Forecasting and validation 18
4. RESULTS 19
4.1 COVID-19 Temporal trends in Bulawayo 19
4.1.1 Mann-Kendall Trend Test.....20
4.1.2 Sen's slope 20
4.2 COVID-19 Spatial clustering and hotspots.....21
4.2.1 Cluster and Outlier analysis 22
4.2.2 Hotspot analysis 23
4.3 COVID-19 Spatiotemporal clustering and hotspots 24

4.3.1 Space-time cube 24
4.3.2 Time series clustering analysis 25
4.3.3 Local Outlier analysis 27
4.3.4 Emerging hotspot analysis 28
4.4 Space-time forecasting of COVID-19 cases 30
4.4.1 Exponential Smoothing vs Forest-based forecasting 30
4.4.2 Evaluating forecasts by location 32
5. Discussion 33
6. Conclusion 35
References 36

LIST OF ACRONYMS

ARIMA	Autoregressive Integrated Moving Average
COVID-19	Coronavirus Disease 2019
ESDA	Exploratory Spatial Data Analysis
GIS	Geographical Information System
GWR	Geographically Weighted Regression
LISA	Local Indicators of Spatial Association
LSTM	Long short-term memory
MGWR	Multi-scale Geographically Weighted Regression
MoHCC	Ministry of Health and Child Care
RF-SSA	Recurrent Forecasting-Singular Spectrum Analysis
RMO	Restricted Movement Order
RMSE	Root Mean Square Error
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SEM	Spatial Error Model
SLM	Spatial Lag Model
STC	Space-Time Cube
WHO	World Health Organization

LIST OF FIGURES

Figure 1- Study Area Map of Bulawayo province, Zimbabwe……. 11

Figure 2- Methodological Flowchart of the research…… 12

Figure 3- Space-Time Cube……..... 14

Figure 4- Graph showing monthly COVID-19 positive cases time series……...19

Figure 5- Graph showing monthly COVID-19 positive cases time series with Sen slope…….. 21

Figure 6- Choropleth map of cumulative positive cases for Apr 2020 to Aug 2021……22

Figure 7- Cluster and Outlier analysis maps……..22

Figure 8- Hotspot analysis maps……...23

Figure 9- Space-time cube visualization……...25

Figure 10- Time series clustering map……..26

Figure 11- Average time series per cluster graph……..27

Figure 12- Local Outlier analysis map……..29

Figure 13- Emerging hotspot analysis map……...30

Figure 14- Exponential smoothing and Forest-based forecasting…….31

Figure 15- Combined forecasting showing evaluation by location……....33

LIST OF TABLES

Table 1- Data description……..12

Table 2- Mann-Kendall results……..20

Table 3- Sen's slope equation……....21

Table 4- Pseudo F-statistic summary……....25

Table 5- Trend statistics for average time series (Mann-Kendall)……..…..26

Table 6- Number of locations per cluster……...27

Table 7- Number of locations per Local outlier analysis category……....28

Table 8- Number of locations per Emerging hotspot analysis category……..…29

Table 9- Forecast RMSE for exponential smoothing vs Forest-based forecasting……..….32

Table 10- Validation RMSE for exponential smoothing vs Forest-based forecasting …….……...32

1. INTRODUCTION

1.1 BACKGROUND

Infectious diseases cause human suffering and they come with public health challenges that cannot be overlooked. The SARS-CoV-2 virus causes Coronavirus Disease (COVID-19), an infectious disease. The first human instances of COVID-19, according to Sun et al., (2020), were discovered in Wuhan, China. The World Health Organization (WHO) declared COVID-19 a public health emergency of worldwide concern on January 30, 2020, and a pandemic on March 11, 2020, after it spread globally in a matter of weeks (Oxford Analytica, 2020).

According to Zimbabwe's Ministry of Health and Childcare (MOH), the country reported 112,435 cases, 81,570 recoveries, and 3,676 fatalities as of August 3, 2021. (Ministry of Health and Child Care (MoHCC) Harare; Zimbabwe, 2021a). The global pandemic has not spared Bulawayo, Zimbabwe's second-largest metropolis. Bulawayo had 11,761 cumulative cases, 9,980 recoveries, and 575 deaths as of August 4, 2021 (Ministry of Health and Child Care (MoHCC) Harare; Zimbabwe, 2021b).

Due to the widespread pandemic, the exacerbation of COVID-19 has drawn increasing global attention. Many countries have enforced nationwide lockdowns with considerable limitations on movement and productivity, sports, entertainment, and education to prevent the spread of COVID-19. According to Chaora (2020), on 27th of March 2020, Zimbabwe was placed under Level 4 lockdown for the first time. Month-long restrictions have been imposed and Level 2 continues for an indefinite period while progress reviews are done occasionally. Following a spike in COVID-19 incidence in mid-July 2020, more limitations were imposed including a dark-to-dawn daily curfew, during which only essential services could operate. Similar but tougher lockdown restrictions were imposed in January 2021 following the second wave of cases and these included a ban in gatherings and a delay in the re-opening of schools.

1.2 RESEARCH PROBLEM STATEMENT

The COVID-19 pandemic has been termed "the greatest challenge that we have faced since the United Nations' creation," (Hall et al., 2020), indicating that it is more than a health issue; it is a crisis already harming economies and communities in a variety of ways. This pandemic has tested global and local emergency management systems, prevention measures, and control mechanisms. According to OCHA (2020), about 80 per cent of COVID-19 cases and deaths came from Harare, Bulawayo, Midlands and Matabeleland South provinces in Zimbabwe with Bulawayo having the second highest. Dams that supplied Bulawayo City were at just 25.6 percent capacity with a deficit of 17 million litres of water

per day for the city's residents (OCHA, 2020). This has contributed to the poor water, sanitation and hygiene situation in Bulawayo leaving residents at high risk of contracting COVID-19. Given the high-risk of COVID-19 in Bulawayo, there is a dire need to monitor and model the spatial trend and spread of infectious diseases that could help in targeted testing, prioritizing vaccination areas, and optimizing the allocation of limited resources. Hence, identification of areas with high-risk for COVID-19 outbreak and the prediction of spatial occurrence of COVID-19 is a major priority of the city council.

1.3 AIM AND OBJECTIVES

1.3.1 Aim

The study aims to map and evaluate the spatial and temporal patterns of COVID-19 and identify the spatiotemporal distribution characteristics and changing trends of cases, using geospatial approaches.

1.3.2 Objectives

The study's primary objectives were to:

1. Identify the COVID-19 temporal trends in Bulawayo

2. Map the geographic distribution of high-risk zones

3. Analyze the spatiotemporal clustering pattern and hotspots of COVID-19 in Bulawayo

4. Predict future COVID-19 spatial distribution based on the spatiotemporal trends

1.4 RESEARCH QUESTIONS

1. Is it viable to model and map the spatiotemporal behavior of COVID-19 incidents using geostatistical techniques?

2. Is it feasible to model and map the high-risk areas of COVID-19 using GIS?

3. Does any clustering or uniformity exist in the study area?

4. Is the prediction of COVID-19 cases possible?

1.5 RESEARCH RATIONALE

GIS-based COVID-19 mapping has been applied successfully in previous studies (Pourghasemi et al. (2020); Rahman et al. (2021); Munanza et al. (2021); (Purwanto et al (2021)).However, the use of geostatistical techniques in the mapping and prediction of COVID-19 cases in Zimbabwe has been limited. According to Gibson et al. (2020), the City of Bulawayo has until recently been limited in its use of GIS to help with challenges such as that which the COVID-19 pandemic presents with a major limitation being the lack of computer hardware and data. In this study, the COVID-19 hotspots in Bulawayo were explored and located using geospatial methodologies and spatial statistics, setting the framework for further geospatial analysis, risk factor evaluation, epidemic monitoring, and management. This study's findings demonstrate the importance of geospatial mapping and spatial statistics, which can swiftly show and build maps by highlighting the geographic and temporal fluctuation of susceptible locations. The hotspot locations were identified, and recommendations were made for crucial preparedness, readiness, and response activities in the face of such a health crisis. Forecasts about where and when the disease will occur may be of great usefulness for decision-makers, as they give them the time to act on the local public health systems and better plan health policy interventions. Forecasting allows policymakers to be proactive and not reactive, which is critical for an effective response. This study will assist government agencies and communities in making effective and well-informed decisions to create and implement COVID-19 control policies in Bulawayo.

2. LITERATURE REVIEW

In Iran, Pourghasemi et al. (2020) conducted research on COVID-19 that included geographical modeling, risk mapping, change detection, and outbreak trend analysis. They used regression modeling, spatial modeling, risk mapping, and change detection utilizing the random forest (RF) machine learning technique and validation of the modeled risk map to forecast mortality trends. A heatmap was used to identify two important time frames for the COVID-19 epidemic in Iran in the study. The regression models for death cases showed an upward trend, but there was evidence of a turning point. The rate of coronavirus infection and the population density of the province were shown to have a polynomial connection. In addition, a third-degree polynomial regression model for deaths lately indicated a growing trend, showing that the outbreak's subsequent efforts were insufficient and unsuccessful.

Lak et al. (2021) investigated 43,000 confirmed COVID-19 cases at the neighborhood level in Tehran, Iran's capital, to better understand the spatial patterns of COVID-19. Using exploratory geographical data analysis and spatial regression, they investigated the spatiotemporal patterns and place-based elements that contributed to the pandemic's spread. A geo-referenced database was created for the study, which included 12 quantitative place-based characteristics linked to physical attributes, land use

and public transit, and demographic status. Local spatial non-stationarity was investigated using a regionally weighted regression model. Population density, dispersion of neighborhood centers, drugstores, and chain stores were discovered to be the key factors contributing to the spread of COVID-19, with population density having the highest R^2. Due to the lack of comparable place-based and socioeconomic data at the neighborhood level in Tehran, the analysis was limited to only a few variables. The study only used data from the first five months of the epidemic in Tehran, which may not be enough to gain a complete picture of the disease's behavior. A more in-depth examination of socioeconomic and environmental factors could help researchers better understand COVID-19 trends and identify possible COVID-19 danger zones.

Cordes & Castro (2020) employed a spatial scan statistic to find clusters of high and low testing rates, high positivity rates, and a high proportion of positives in New York City. To determine relationships between outcomes, clusters, and contextual factors, they employed boxplots and Pearson correlations. Clusters with lower testing rates and a lower proportion of positive tests had higher wealth, education, and a white population, whereas clusters with high testing rates and a high proportion of positive tests were disproportionately black and lacked health insurance. With proportion positive tests, correlations revealed inverse connections of the white race, education, and income, and positive associations with the black race, Hispanic ethnicity, and poverty. They suggested directing testing and health-care resources to eastern Brooklyn, which had minimal testing and a high proportion of positives. Due to the great spatial resolution, the study examined zip code level data, which provided a precise insight into the COVID-19 neighborhoods. However, the study simply looks at correlations between COVID-19 testing patterns and zip code contextual factors; there is no causal relationship between the variables.

Using a 154-day real-time epidemiological data series, Rahman et al. (2021) assessed the COVID-19 outbreak and transmission dynamics in Bangladesh over time and space. For the modeling of COVID-19 vulnerability zoning at the district level, they used Getis-Ord Gi* statistics and the analytical hierarchy process-based weighted sum technique. COVID-19 spreads from the center, according to the spatial analysis, with Dhaka being the most exposed district, followed by Chattogram, Narayanganj, Cumilla, and Bogra. In both the cases and deaths distributions, a single significant clustering pattern in the central region, which extended out predominantly to the south-eastern part, was identified as a prime hotspot. In addition, potential links between illness transmission and selected factors that aid in disease spread were discovered. The western, southwestern, north-western, and north-eastern areas were designated as highly vulnerable zones, while the central, eastern, and south-eastern parts were designated as medium susceptible zones. The temporal analysis revealed that in a densely populated country, the number of tests performed per day was insufficient, and that as the number of tests

increased, so did the number of positive cases, and vice versa; thus, increasing the number of tests can be an effective measure to obtain the true picture.

Munanza et al. (2021) undertook a global analysis to examine COVID-19 incidence rates and locate hotspots and clusters. They calculated monthly incidence rates for each nation using monthly data on COVID-19 cases from the World Health Organization dashboard. They looked at the spatial clustering of Covid-19 occurrence in all months of the year using the Global Moran's I and Anselin Local Moran's I. The Getis-Ord G* was then used to assess the data's hotspot and cold spot variations. From January to April, there was a movement in COVID-19 hotspots and significant clustering from the Western Pacific to Europe and the Americas. From July, Eastern Mediterranean countries became disease hotspots, leaving Africa as a cold spot for the entire year. In December, there was a lot of clustering, whereas, in June, there was a lot of clustering. One of the study's key flaws is that it was conducted on a worldwide scale rather than on microscopic data, which can lead to considerable results distortions due to age standardization and varied population numbers.

Arashi et al. (2020) used heatmap retrieval of hotspot areas and spatial analysis with the Moran index to investigate the spread of COVID-19. The generalized logistic growth curve model was used to forecast COVID-19 distribution and identify hotspots for virus management. The findings revealed that in South Africa, the spreading direction is primarily determined by proximity to neighbors, regardless of distance from the infectious center. The logistic growth models showed a good fit to the provincial data, but a different modeling method for Gauteng, Free State, and Northern Cape provinces could provide even better results.

Ramírez-Aldana et al. (2020) investigated how COVID-19 cases are distributed regionally and how socioeconomic and climatic characteristics of Iranian provinces may influence the number of cases. To derive substantial clustering, they used a Bayesian approach using the global (Moran's I) and local markers of spatial autocorrelation (LISA). They also used a multivariate spatial lag model to better explain COVID-19 incidence by including a set of variables that could influence the disease's existence. Around Tehran, there was a high concentration of COVID-19 patients. COVID-19 was also revealed to be sensitive in urbanized provinces with older population structures and higher temperatures. They also discovered that literacy is strongly linked to COVID-19.

Using spatial autocorrelation, hot spots, and Spatio-temporal scanning statistics, Wang et al. (2021) investigated the Spatio-temporal behavior of COVID-19 transmission and its affecting factors in China from January to October 2020. The association between contributing factors and verified COVID-19 cases at the provincial level was investigated using Spearman rank correlation analysis and multiple

linear regression. COVID-19's distribution in China appears to follow time, with spatial correlation and notable clustering locations, according to the study. Spatiotemporal scanning analysis revealed that the most COVID-19 high incidence months occurred in the early months of the epidemic, from January to March, and Hubei had the highest aggregation risk. It was concluded that COVID-19 transmission was influenced by environmental temperature, and that the migration index has a stimulating effect on COVID-19 transmission.

Mo et al. (2020) employed the space-time cube model to identify the space-time clustering pattern, determine cold and hot spot trends, and display the visual output for the COVID-19 epidemic in China from January 23 to February 24, 2020. The space-time cube-based spatiotemporal analysis technology can analyze the spatiotemporal pattern of epidemiology data in-depth and provide a visual output of the results that can intuitively portray the distribution and trend of data in space-time. According to the report, the pandemic spread quickly across the country in a short period. It also revealed that COVID-19 cases were distributed in an inconsistent spatiotemporal pattern. Wuhan and Shiyan cities were in the core of the spatiotemporal clustering pattern, with both cities having a high-high clustering pattern with a surrounding unstable multiple-type pattern in partial parts of Henan, Anhui, Jiangxi, and Hunan provinces, and Chongqing city. Because timely models play a significant part in making health decisions, the findings of such a study may offer decision-makers basic information. The study used city-level data, which may have resulted in numerous spatial data points being overlooked, lowering the accuracy of the spatial analysis.

Purwanto et al. (2021) conducted a spatial analysis using the Emerging hotspot Analysis and the Space-Time Cubes over four months in East Java, Indonesia. COVID-19 distribution in East Java was focused in Surabaya, then spread predominantly to suburban areas and other cities, according to the findings. Oscillating patterns and sporadic hotspots plagued both cities. The pattern demonstrated that newly infected patients always follow the recovery of infected COVID-19 patients and that the number of positive patients grows quicker than the number of recovered patients. The monthly hotspot study's findings revealed a precise COVID-19 spatiotemporal pattern, allowing for a more in-depth examination of events and policies in terms of location and time.

Ngwira et al. (2021) investigated the distribution of COVID-19 in Malawi in terms of both space and time. With time and district as independent variables, and district as the spatial unit, semi-parametric spatial-temporal models were fitted to the number of monthly confirmed cases as outcome data, accounting for sociodemographic aspects. Location and time have a significant impact on the study's findings, with the two interacting. COVID-19 was more prevalent in urban regions than rural areas. COVID-19 risk fluctuated over time in most districts, with rural areas always being at a lower risk.

According to the study, initiatives targeted at preventing the spread of COVID-19 should target big cities, with a focus on periods that have shown significant risk, and a focus on the elderly and wealthy.

The Mann-Kendall test was employed by Shaharudin et al. (2021) to examine the trend of the COVID-19 outbreak in Malaysia. They also measured and predicted daily COVID-19 cases in Malaysia for 10 days using verified cases and the Recurrent Forecasting-Singular Spectrum Analysis (RF-SSA) model. The early Mann-Kendall test findings showed a diminishing trend pattern for new cases during RMO (Restricted Movement Order) 3 compared to RMO 1, RMO 2, and RMO 4, with a major spike in the COVID-19 epidemic during RMO1. The RF-SSA over-forecasted the instances by 0.36 percent, demonstrating the model's high accuracy. Malaysia would reach single digits in daily confirmed COVID-19 cases by early June 2020, according to the recommended model. The data demonstrated that the RF-SSA model can detect trends and accurately predict COVID-19 instances. However, it was determined that the RF-SSA method needs to be improved to capture extreme data changes more effectively.

Galasso et al. (2021) utilized a non-parametric random forest model to forecast COVID-19 cases at the county level in the United States. For each reference week between 11/01/2020 and 01/10/2021, the model was used to project the first 4 weeks for 3068 counties. The random forest model continuously beat the gold-standard compartmental models, with a mean absolute error of fewer than 300 weekly cases/100,000. The non-parametric random forest model has a lot of potential in ensemble modeling since it can handle a larger training feature set while still maintaining good performance and using few resources.

Barman (2020) investigated the performance of different Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) models in estimating the number of confirmed COVID-19 cases across several nations. Several LSTM models and ARIMA were used to generate 1-day, 3-day, and 5-day forecasts utilizing time series of daily cumulative COVID-19 cases. When compared to ARIMA models, some observed that LSTM models had the lowest prediction errors. COVID-19 cases were slightly underestimated by the LSTM models, whereas ARIMA forecasts were slightly inflated by ARIMA. While ARIMA requires larger sequences, it was revealed that LSTMs may function reasonably well with sequence lengths as short as three. LSTMs necessitates many training samples. Both LSTMs and ARIMA were useful for time series analysis and forecasting for COVID-19 based on the proposed k-period performance metrics.

In Gorontalo, Indonesia, Djakaria & Saleh (2021) investigated the COVID-19 prediction model utilizing Holt-Winters exponential smoothing. COVID-19 data was used in this study, which ran from

April 10 through October 13, 2020. The study's major goal was to identify the best forecasting model with the right smoothing settings. The best forecasting model with smoothing values of = 0.1 and = 0.5 for trend and seasonality, respectively, was discovered to have the minimum MAPE value of 6.14 when utilizing the Holt-Winters exponential smoothing.

Yang et al. (2020) used the Mann–Kendall and Pettitt methods to find COVID-19 spatial and temporal patterns in Hubei province, China. The Local Moran's I index was used to discover the geographic patterns of the incidence rate, including spatial clusters and outliers. The improvement in the capacity to diagnose the condition was due to the pattern, which was characterized by a growing trend and abrupt shift. Due to the high intensity of population spatial interaction, geographical clusters with high incidence rates over the period were centered in the Wuhan Metropolitan Area.

Kim & Castro (2020) conducted research to determine how COVID-19 was distributed among South Korean districts and whether the spatial distribution and duration of clusters changed as a result of the country's lockdown strategy. From January 20 to May 31, 2020, a spatiotemporal study of COVID-19 daily confirmed cases by 250 districts in South Korea was conducted. To detect spatial autocorrelation, they used the global Moran's I. COVID-19 spatiotemporal clusters were then analyzed using the space-time scan statistic. The geographical pattern exhibited high spatial autocorrelation in the global Moran's I. By using the space-time scan statistic, a discrete Poisson model revealed twelve statistically significant spatiotemporal clusters. Clusters' spatial layout changed over time, and the lifespan of clusters shrank. The findings showed that South Korea's COVID-19 containment policy was very effective in both early identification and mitigation, with recent clusters being small and short-lived. Due to data availability for the space-time cluster analysis, daily confirmed cases for some districts were aggregated at the city level, which constituted a study restriction.

To better understand how various variables could have caused COVID-19 distribution at the district level in India, Dutta et al. (2021) employed GIS-based spatial models. They used 19 variables to explain the disease's variability. To investigate spatial dependency, spatial lag and error models (SLM and SEM) were used, while geographical weighted regression (GWR) and multi-scale GWR (MGWR) were used to investigate at the local level. It was observed that global models do not do a good job of describing the elements that cause COVID-19 outbreaks. With the lowest AICc value, MGWR was found to be the best-fit model for explaining the variables affecting COVID-19. COVID-19 infections were shown to be especially sensitive in areas with high population density, urbanization, and bank facilities. These findings highlighted the importance of effective strategies addressing social isolation and limited mobility. There was no evidence of a link between COVID-19 and the climate in the study. The study's shortcoming is that it did not account for a variety of pre-existing underlying health

conditions that could contribute to COVID-19. The incorporation of underlying health concerns, according to Bansal (2020), may help influence COVID-19.

Mollalo et al. (2020) examined at COVID-19 incidence at the county level in the United States. They explored 35 environmental, socioeconomic, topographic, and demographic variables to see if they could explain COVID-19's regional variation. They investigated spatial dependence using SLM and SEM, as well as locally examining spatial non-stationarity using GWR and MGWR models. The results showed that these models still performed poorly when compared to local models. Furthermore, as compared to the others, MGWR could explain the highest variability with the lowest AICc. Modeling different lockdown policies and the degree at which such restrictions were adopted and enforced was a study constraint. This is because different states had varied policies and procedures, and these policies and their implementations might have a huge impact on COVID-19 incidence rates.

Vahedi et al. (2020) used spatiotemporal lags of infection rates, human interactions, human mobility, and county socioeconomic composition as predictive features to develop a Spatiotemporal autoregressive model to predict county-level new cases of COVID-19 in the coterminous US. They employed Facebook and cell phone-derived metrics of connection and human mobility to capture human interactions and used them in two independent models to predict new COVID-19 cases at the county level. The model was tested on 14 forecast dates ranging from one to four weeks in length. When compared to a baseline model generated by the COVID-19 Forecast Hub, their estimates improved by an average of 6.46 percent. The Mean Absolute Errors across the two-week prediction horizon improved by 20.22 percent in the four-week prediction horizon, indicating that the model has great predictive power over longer prediction horizons.

Based on climatic data provided by bioclimatic variables, Bariotakis et al. (2020) predicted the global expansion of COVID-19 infections. The estimated probability maps revealed that infection risk areas migrate from tropical to temperate and Mediterranean bioclimatic zones. Infection rates increased in Europe, then spread to areas of the Middle East and Northern Africa, as well as the eastern coasts of North America, the south-eastern coasts of Latin America, and two areas of Southern Australia, according to maps. The method could be useful in incorporating climatic factors into the development and implementation of public health policy.

After accounting for community-level factors such as demographics, socioeconomic status, and human mobility status, Wang et al. (2020) used a retrospective cross-sectional regression analysis using the Fama-MacBeth procedure to assess the impact of temperature and relative humidity on COVID-19 transmission across communities. They used data from 100 Chinese cities with COVID-19 daily

symptom-onset cases and 1005 US counties with COVID-19 daily confirmed cases. The effective reproduction numbers were calculated using a total of 69 498 instances in China and 740 843 cases in the United States. In both China and the United States, statistically significant negative associations between temperature/relative humidity and the effective reproductive number (R-value) were discovered. COVID-19 transmission may be inhibited by greater temperatures and relative humidity, according to the findings. The study suggests that the summer and rainy seasons may reduce COVID-19 transmissibility, but the COVID-19 pandemic is unlikely to 'automatically' diminish in the summer. In some parts of the Northern Hemisphere, cold and dry seasons have the potential to disrupt the delicate transmission balance and weaken downward trends. Their findings, however, are unable to confirm the precise processes underlying the relationship between temperature and COVID-19 transmissibility.

Sun et al. (2020) employed a geographic regression model, spatial lag, spatial error, and spatial autoregressive combination models to investigate how spatial features, such as county linkages, affect COVID-19 prevalence in the United States. The aspatial ordinary least squares regression model was discovered to overestimate COVID-19 period prevalence in counties with low observed rates, however spatial modeling can successfully overcome this issue. The period prevalence for counties was better predicted using spatial models, especially along the Atlantic coasts and through the Black Belt. Overall, the model fit was good, with little variation, across counties on both coasts. They concluded that spatial models could help explain some of the geographic differences in COVID-19 prevalence. These models highlight spatial heterogeneity in model fit, allowing researchers to pinpoint sections of the country where the fit is heterogeneous and warrants further investigation in the near term.

Despite numerous efforts to map and assess COVID-19 incidence, more research on the pandemic is still needed. Localized research may be more useful in supporting local governments in making well-informed decisions. Because more data is available now, the epidemic may be studied over longer periods. Geospatial methodologies and spatial statistics could be used in Bulawayo, setting the framework for further geospatial analysis, risk factor evaluation, epidemic monitoring, and management. These techniques can be used to swiftly show and build maps by highlighting the geographic and temporal fluctuation of susceptible locations. Forecasts about where and when the disease will occur may be of great usefulness for decision-makers, as they give them the time to act on the local public health systems and better plan health policy interventions.

3. MATERIALS AND METHODS

3.1 Study area

The study area for the research is Bulawayo, a city in the southwestern part of Zimbabwe. Bulawayo is found at 20.1378° S and 28.56345° E. It is Zimbabwe's second-largest city covering an area of 1,706 sq. kilometers and has a population of 639,722 (World Population Review, 2022). Bulawayo comprises of four administrative districts which are Imbizo, Khami, Mzilikazi, and Reigate, and it is divided into 29 wards, as depicted in Figure 1.

Bulawayo has a subtropical climate due to its relatively high altitude. It has a mean annual rainfall of between 580mm and 610 mm, with most rain falling during the summer from December to February. Temperatures in Bulawayo range from an average of 21° C in July to 30° C in October.

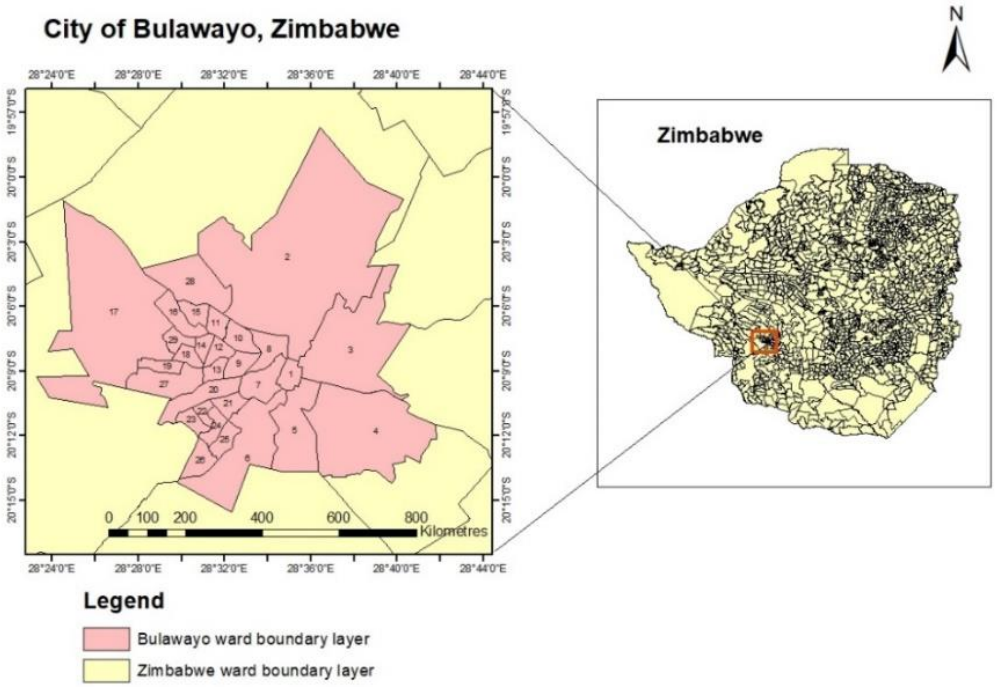


Figure 1: Study Area Map of Bulawayo province, Zimbabwe

3.2 Research design

A brief structure of the research design is shown in Figure 2 below. It shows the flow and interconnection between the inputs, processes/methods applied, and the outputs of the research.

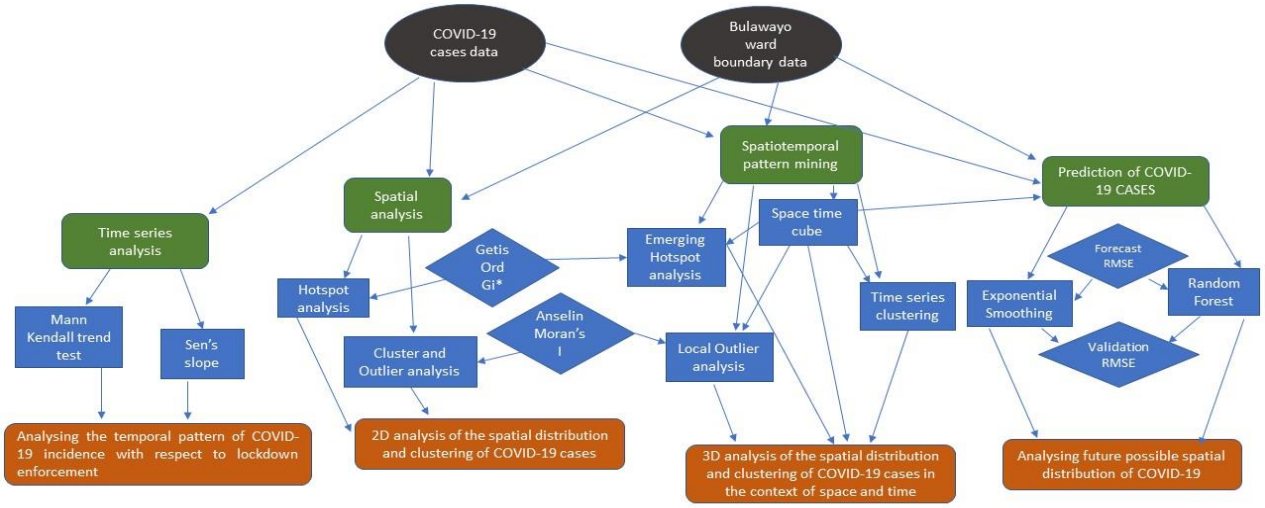


Figure 2: Methodological Flowchart of the research

3.3 Data requirement

The research used ward level data of all confirmed COVID-19 cases reported between April 2020 and August 2021 acquired from the City of Bulawayo Municipality. The health services department at the City of Bulawayo receives all daily COVID-19 cases data from testing centers and hospitals. A confirmed COVID-19 case, according to the Zimbabwean Ministry of Health, is defined as someone who has had laboratory confirmation of the virus with clinical signs and symptoms, and who has had a Reverse Transcription Polymerase Chain Reaction (RT-PCR) test positive for SARS-CoV-2. Table 1 provides a full overview of the dataset.

Table 1: Data description

Data set	**Factors**	**Source**	**Temporal description**
COVID-19 data	Total cases, total deaths, total recoveries	City of Bulawayo Municipality	Monthly data April 2020 to- August 2021 at ward level of Bulawayo city

3.4 Methodology

3.4.1 Mann- Kendall statistic

The Mann-Kendall Trend Test was used to determine if the COVID-19 cases pattern was continuously increasing or decreasing. Mann (1945) was the first to suggest the test, and Kendall (1975) examined it further. Hirsch & Slack (1984) then improved it by factoring in seasonality. According to Mann (1945), the test is based on two hypotheses: the null hypothesis (H0) stating that there is no trend, while the alternate hypothesis (H1) states there is a significant upward or downward trend. The data did not have to meet the normality assumption because the test is non-parametric.

To determine trends, the number of COVID-19 positive instances was examined as an ordered time series, with each case's value compared to all subsequent values. For a time series $x_1, \ldots, x_n$, the Mann-Kendall test uses the following statistic:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign(x_j - x_k) \quad (1)$$

$$\text{Sign } (x_j - x_k) = \begin{cases} +1, if\ (x_j - x_k) > 0 \\ 0, if\ (x_j - x_k) = 0 \\ -1, if\ (x_j - x_k) < 0 \end{cases} \quad (2)$$

where n is the length of the sample, x_k and x_j represent the counts of COVID-19 cases at time k and j where k=1, 2, …, $n-1$ and $j = k+1$, …, n (Meals et al., 2011).

The Mann-Kendall statistic, S, was initially assumed to have a value of zero representing the absence of a trend. If COVID-19 cases for a later month were higher than those of the earlier month, S was increased by one. If COVID-19 cases for a later month were lower than those of the earlier month, S was decremented by one, as illustrated in equation 2. As stated in equation 1, the final value of S was the sum of all such increments and decrements. An upward trend was indicated by a high positive S value, whereas a downward trend was indicated by a low negative S value. However, to examine the trend scientifically, the probability associated with S and the sample size, n, was determined. If the p-value was less than 0.05, we rejected the null hypothesis, which means there was no trend in the data, using a 5% significance level. If the p-value was greater than 0.05, however, the null hypothesis was not rejected.

3.4.2 Sen slope estimator

The magnitude of trends in the COVID-19 positive cases data was calculated using Sen's slope estimator. Developed by Sen (1968), the Sen's slope estimator was calculated in the same way as the Mann-Kendall statistic, but instead of combining all signs of the pairs, it calculated slopes for all pairs

of all ordinal time points and then determined the median of all the slopes to indicate the magnitude of the trend in equation 3 below. For the set of pairs (i, x_i) where x_i is a time series, Sen's slope was defined as:

$$\text{Sen's slope} = \text{Median}\left\{\frac{x_j - x_i}{j - i}\right\} ; i < j \} \quad (3)$$

where x_i and x_j represent the counts of positive cases at the time i and j, respectively (Meals et al., 2011).

3.4.3 Space-Time cube

A Space-Time Cube (STC) is a tool that can study how phenomena evolve in geographic space. Using time series analysis, integrated spatial and temporal pattern analysis, and advanced 2D and 3D visualization techniques, a space-time cube can visualize and analyze spatiotemporal data (Kveladze et al., 2018). COVID-19 positive cases were used as station data where geography did not change, but cases changed over time into space-time bins. Each cube in a space-time cube represented a slice of time, as seen in Figure 3. A temporal change in the geographic area could be seen by looking at the slices of time from bottom to top.

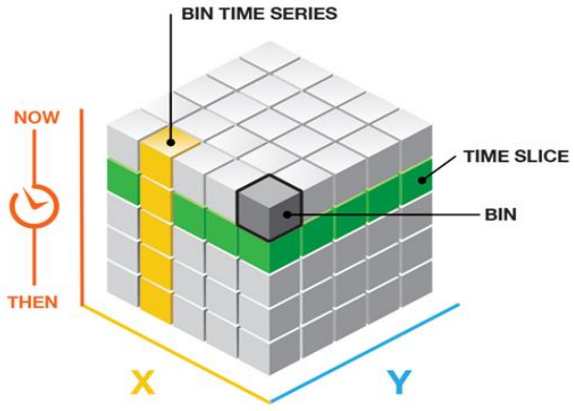


Figure 3: Space-Time Cube (Kang et al. 2018)

A z-score and a p-value represented the trend for each bin time series. A trend with a low p-value was statistically significant. The sign associated with the z-score determined whether the trend was an increase or a decrease in bin values (Kveladze et al., 2018).

3.4.4 Anselin Local Moran's I

According to Hongfei et al. (2007), the Local Moran's I statistic is a local spatial autocorrelation statistic that identifies local clusters or outliers to better understand their contribution to the 'global' clustering statistic. Anselin (1995) coined the term "Local Indicators of Spatial Association" to describe a group of local indicators (LISAs). The Local Moran's I statistic decomposes the global Moran's I statistic into the degree of spatial connection associated with each observation, providing insight into data behavior at local levels. In exploratory spatial data analysis (ESDA), LISAs serve two purposes: they discover local spatial clusters, and they undertake sensitivity analysis (identify outliers). Anselin's Local Moran's I was utilized to provide information on high-risk and low-risk areas of COVID-19 at the ward level in Bulawayo. High-high, low-low (showing a clustering pattern), low-high, high-low (indicating an outlier pattern), and not significant were the five categories this index separated wards into. High-high (HH) identified places with a high incidence of disease, and their surrounds, and hence identified high-risk clusters or hotspots. Low-low (LL) spots depicted locations where COVID-19 incidence was low and represented low-risk disease clusters or cold spots. The Local Moran's I statistic of spatial association is given as:

$$I_i = \frac{x_i - \bar{X}}{{S_i}^2} \sum_{j=1, j \neq i}^{n} w_{i,j}\left(x_j - \bar{X}\right) \quad (4)$$

Where x_i is an attribute for feature i, $\bar{X}$ is the mean of the corresponding attribute, $w_{i,j}$ is the spatial weight between feature I and j and:

$${S_i}^2 = \frac{\sum_{j=1, j \neq i}^{n} \left(x_j - \bar{X}\right)^2}{n-1} \quad (5)$$

With n equating to the total of features.

3.4.5 Getis-Ord Gi*

The Getis-Ord Gi* was used to identify clusters of COVID-19 incidence within Bulawayo. The clusters reflected either high or low values of COVID-19, corresponding to hot and cold areas, respectively.

The Getis-Ord Gi* statistic is calculated by comparing the sum of all ward level COVID-19 cases to the local ward level COVID-19 cases data and those of its neighbors.

A hotspot was identified as a geographic area with a higher concentration of COVID-19 incidences than the expected number of random distributions. A high z-score and low p-value implied a significant hotspot, while a negative and low z-score indicated a significant cold spot. Getis-Ord Gi* was utilized in this work to distinguish high-value clusters (hotspot) from low-value clusters (cold spots). According to Getis & Ord (1992), the Getis-Ord Gi* statistics is expressed as:

$$G_i * = \frac{\sum_{j=1}^{n} w_{i,j} w_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}{}^2 - \left(\sum_{i,j}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \quad (6)$$

where xj is the attribute value for feature j; wi,j is the spatial weight between feature i and j; n is equal to the total of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (7)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j{}^2}{n} - (\bar{X})^2} \quad (8)$$

3.4.6 Holt-Winters (Triple exponential) smoothing

Holt-Winters smoothing is a time series forecasting technique that was used to model three characteristics of the COVID-19 time series: a typical value (average), a slope (trend) through time, and a cyclical repeating pattern (Holt, 2004). It is also known as triple exponential smoothing. The approach is named after Charles Holt and Peter Winters, two people who helped develop it (Winters, 1960).

The Holt-Winters approach used exponential smoothing to encode historical COVID-19 incidence values and used them to forecast the present and future values of COVID-19. Using an exponentially weighted moving average to smooth a time series is known as exponential smoothing.

The Holt-Winters exponential smoothing method was used in space and time to deconstruct COVID-19 cases time series at each ward in Bulawayo of a space-time cube into seasonal and trend components to properly estimate future time steps at each location. A map of the last anticipated time step was the principal result. The components of Winters' Triple exponential smoothing are shown in the following formulae:

$$s_0 = x_0 \quad (9)$$

$$s_t = \alpha(x_t - c_{t-L}) + (1-\alpha)(s_{t-1} + b_{t-1}) \quad (10)$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1} \quad (11)$$

$$c_t = \gamma\frac{x_t}{s_t} + (1-\gamma)c_{t-L} \quad (12)$$

where S_t is the smoothed statistic, it is the simple weighted average of recent observation x_1; $S_{(t-1)}$ is the previous smoothed statistic; α is the smoothing factor of data such that $0 < \alpha < 1$; t is the period; b_t is the best estimate of the trend at time t; β is the trend smoothing factor such that $0 < \beta < 1$, c_t is the sequence of seasonal correction factor when t; and γ is the seasonal change smoothing factor such that $0 < \gamma < 1$ (Winters, 1960).

Triple exponential smoothing is the most advanced kind of exponential smoothing, and it can also produce double and single exponential smoothing models through modification. Holt-exponential Winter's smoothing is an adaptable approach that permits the level, trend, and seasonality patterns to alter.

3.4.7 Random forests

According to Breiman (2001), random forests are a supervised machine learning technique for classification and regression that works by building many decision trees during training. For classification tasks, the random forest's output is the class chosen by most trees. The mean or average prediction of the individual trees is returned for regression tasks. Each tree generates its prediction, which is then utilized to construct final predictions as part of a voting mechanism. The final projections are based on the entire forest rather than a single tree. Using the complete forest rather than a single tree, and the inclusion of a random subset of the training data and a random subset of explanatory variables in each tree that makes up the forest, help avoid overfitting the model to the training dataset.

Tin Kam Ho (1995) developed the first random decision forest method, which was later improved by Leo Breiman (2001). The extension combines Breiman's "bagging" concept with Ho's initial introduction of random feature selection to create a set of decision trees with controlled variance. The bagging approach is used in the training process for random forests. Given a training set $X = x_1, ..., x_n$ with responses $Y = y_1, ..., y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to these samples:

For $b = 1, ..., B$:

1. Sample, with replacement, n training examples from X, Y; call these X_b, Y_b.

2. Train a classification or regression tree f_b on X_b, Y_b.

After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x':

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \tag{13}$$

or by taking the majority vote with classification trees.

The space-time cube generated for Bulawayo COVID-19 cases from April 2020 to August 2021 was used as input for generating the random forests. The training data used to build the forest regression model was constructed using time windows on each location of the space-time cube. To forecast future COVID-19 values for each ward, the forest model was trained by associating the number of cases and the temporal component for each location. Time windows were constructed in which the time steps inside each time window were used as explanatory variables, and the next time step after the time window was the dependent variable to create multiple sets of explanatory and dependent variables within a single time series. Using the trained forest model, the final time steps of the location were used as explanatory variables to forecast the first future time step, which was COVID-19 cases for the month of September 2021.

3.4.8 Forecasting and validation

While forecasting each time series, two models were built for each ward in Bulawayo. The forecast model was the first, and it was used to predict COVID-19 cases of future time steps. The validation model was used to verify the predicted COVID-19 cases.

The forecast model was built by forecasting each position of the space-time cube's time series values. After then, the model was utilized to forecast future time steps. The forecast root mean square error (RMSE), which is equal to the square root of the average squared difference between the exponential smoothing model and the values of the time series, was used to assess the model's fit to each time series.

$$Forecast\ RMSE = \sqrt{\frac{\sum_{t=1}^{T}(c_t - r_t)^2}{T}} \tag{14}$$

where T is the number of time steps, c_t is the fitted value from exponential smoothing, and r_t is the raw value of the time series at time t.

The validation model was used to see how successfully each time series's forecast model could predict COVID-19 future values. It was made by removing some of each time series's last time steps and fitting the exponential smoothing model to the data that was not removed. This model was then used to forecast the hidden data's values, and the forecasted values were compared to the hidden data's raw values. By default, 10% of the time steps were withheld for validation, but the number of Time Steps to Exclude for Validation option could be modified. The Validation RMSE metric, which is equal to the square root of the average squared difference between the forecasted and raw values of the omitted time steps, was used to assess forecast accuracy.

$$Validation\ RMSE\ = \sqrt{\frac{\sum_{t=T-m+1}^{T}(c_t - r_t)^2}{m}} \qquad (15)$$

where T is the number of time steps, m is the number of time steps withheld for validation, c_t is the value forecasted from the first T-m time steps, and r_t is the raw value of the time series withheld for validation at time t.

4. RESULTS

4.1 COVID-19 Temporal trends in Bulawayo

This section provides a temporal overview of COVID-19 positive cases recorded from April 2020 to August 2021 for the City of Bulawayo. Figure 4 below represents COVID-19 cases in Bulawayo over time.

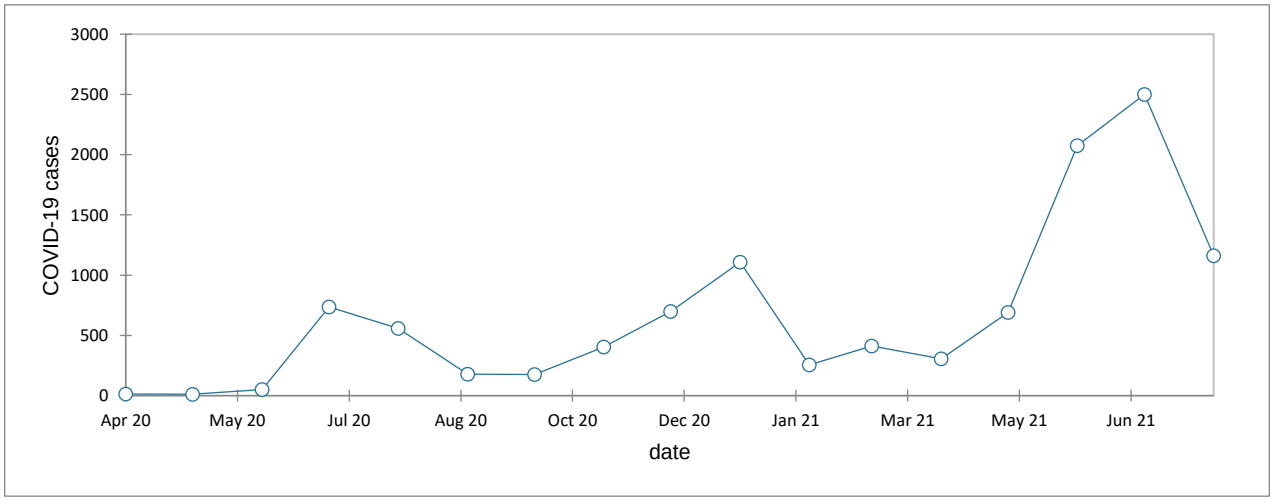


Figure 4: Graph showing monthly COVID-19 positive cases time series

A low number of cases was identified at the beginning of the period as there were 14 cases recorded in April and 12 in May. Positive cases of COVID-19 rose as July recorded 736 positive cases. According to Murewanhema et al., 2020, this period was called the first wave in Zimbabwe. The rise of COVID-19 cases in July marked the peak of the first wave of COVID-19 cases in Bulawayo. Due to the spike in COVID-19 cases, the president announced a dusk-to-dawn curfew, during which only essential services could operate. A decline in cases was identified from July to August 2021.

A gentle rise in COVID-19 cases was noted from October 2020 which recorded 175 cases to a peak of 1107 cases recorded in January 2021. A decline in monthly cases was noted from February which recorded 256 cases. According to Murewanhema et al., 2020, this period was regarded as the second wave.

In March 2021, 411 cases were recorded and a steep rise in positive cases was identified. 2500 cases were recorded in June 2021, marking the beginning of the third wave of COVID-19 cases.

4.1.1 Mann-Kendall Trend Test

A Mann-Kendall Trend Test was run on the monthly COVID-19 cases data to detect monotonic trends in the COVID-19 time series. Table 2 below represents the results from Mann- Kendall trend test. Due to a high positive value of S = 76 and a significance value of 0.002, it was concluded that there was an upward (increasing) trend in the monthly COVID-19 cases in Bulawayo.

Table 2: Mann-Kendall results

Kendall's tau	0,559
S	76,000
V Var(S)	589,333
p-value (Two-tailed)	0,002
Alpha	0,050

4.1.2 Sen’s slope

Sen’s slope estimator was applied to the monthly temporal data to determine the magnitude of the increasing trend identified in the COVID-19 positive cases data. Table 3 shows the results of the Sen, slope estimator.

Table 3: Sen's slope equation

	Value	Lower bound (95%)	Upper bound (95%)
Slope	2,118	0,875	4,930
Intercept	-93089,632	-217034,252	-38199,531

Using Sen's slope test, we found that over the 17 months, the slope was 2.118 (95% CI: 0.875 to 4.93). This shows a gently increasing trend of COVID-19 cases in Bulawayo. Figure 5 shows a depiction of the Sen slope over the COVID-19 time series.

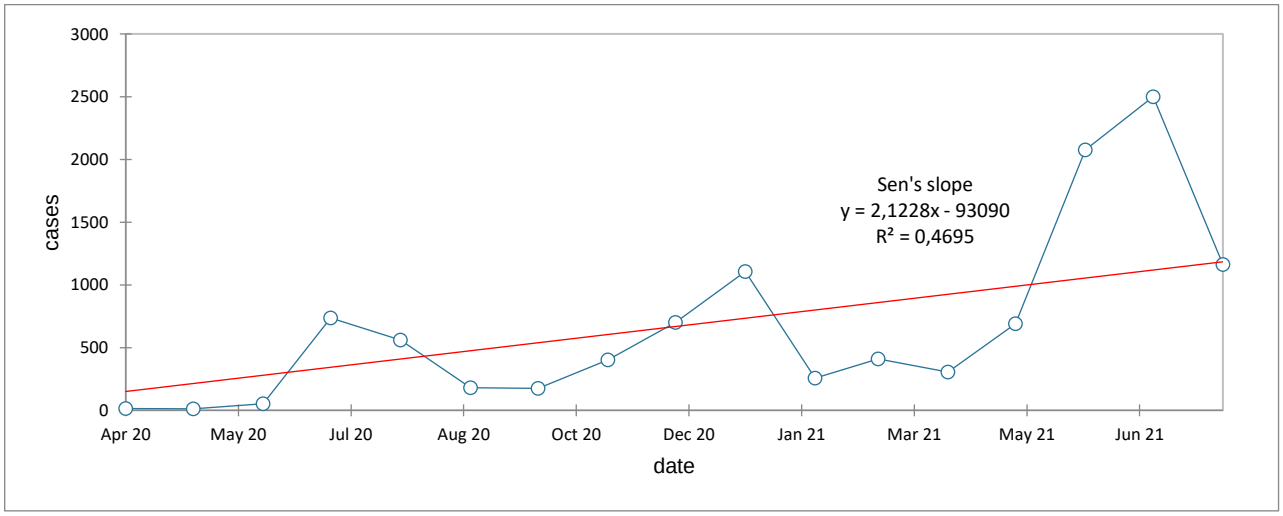


Figure 5: Graph showing monthly COVID-19 positive cases time series with Sen slope

4.2 COVID-19 Spatial clustering and hotspots

This section generally inspects the ward level spatial distribution of the total confirmed cases in Bulawayo. Figure 6 is a choropleth map of cumulative COVID-19 positive cases from April 2020 to August 2021. Majority of cases were found in the eastern parts of Bulawayo with ward 3 having the highest cumulative cases ranging between 735 to 1142 cases. The least cumulative cases were recorded in wards 14, 15, and 17 with cases ranging between 90 to 122 cases. Most wards found in the central parts of Bulawayo had cases ranging between 123 and 700 cumulative cases.

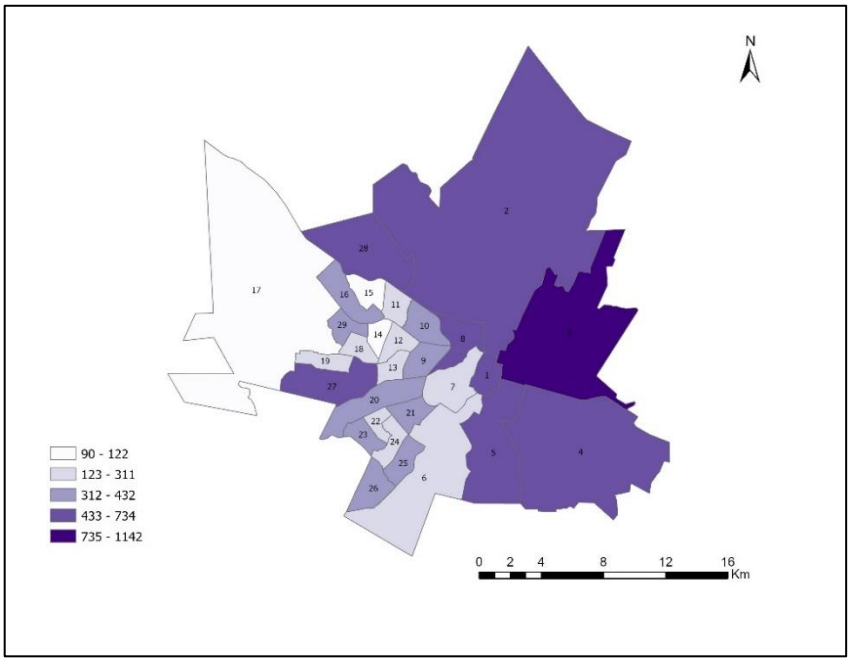


Figure 6: Choropleth map of cumulative positive cases for Apr 2020 to Aug 2021

4.2.1 Cluster and Outlier analysis

Results from the temporal trend revealed three waves of COVID-19 cases in Bulawayo during April 2020 to August 2021 period. Of concern was the spatial distribution of the disease during the identified COVID-19 waves. A Cluster and Outlier analysis was done for COVID-19 cumulative cases for the first, second, and third wave of the disease in Bulawayo using the Anselin Moran's I (Figure 7).

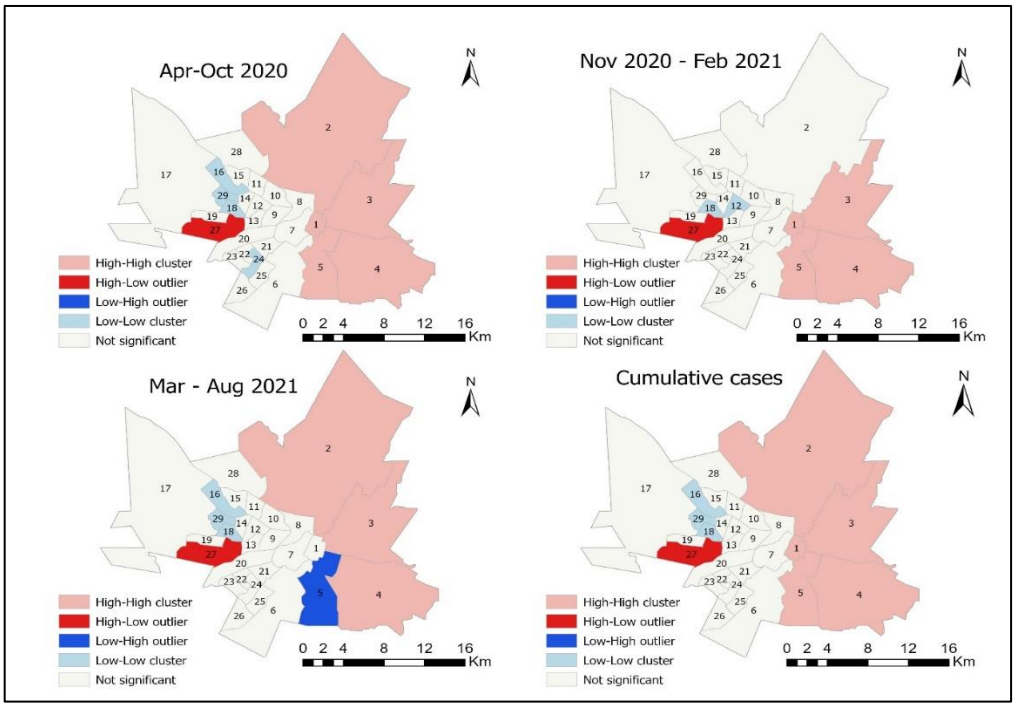


Figure 7: Cluster and Outlier analysis maps

The Figure above shows spatial clusters and outliers of COVID-19 cases within Bulawayo from April 2020 to August 2021. Wards 1, 2, 3, 4, and 5 were identified as regions of high-high clusters throughout the period. However, ward 2 was not significant in the second wave and ward 5 was identified as a low-high cluster in the third wave.

In the first wave, wards 16, 18, 24, and 29 were identified as low-low clusters and in the second wave, only ward 18 and 12 were identified as low-low clusters. Wards 16, 18, and 29 were identified as low-low clusters in the third wave following a similar pattern as that of the first wave. Ward 27 was identified as a high-low outlier throughout the whole period.

4.2.2 Hotspot analysis

Figure 8 shows the results of the Getis-Ord G* hotspot analyzes for the first three waves of COVID-19 from April 2020 to August 2021. Hotspots were identified in the eastern parts of Bulawayo for the first wave, with wards 3, 4, and 5 being the most significant hotspots with 99% confidence. Wards 14, 16, 18, 24, and 29 were identified as significant cold spots, with ward 18 being the most significant cold spot with 95% confidence.

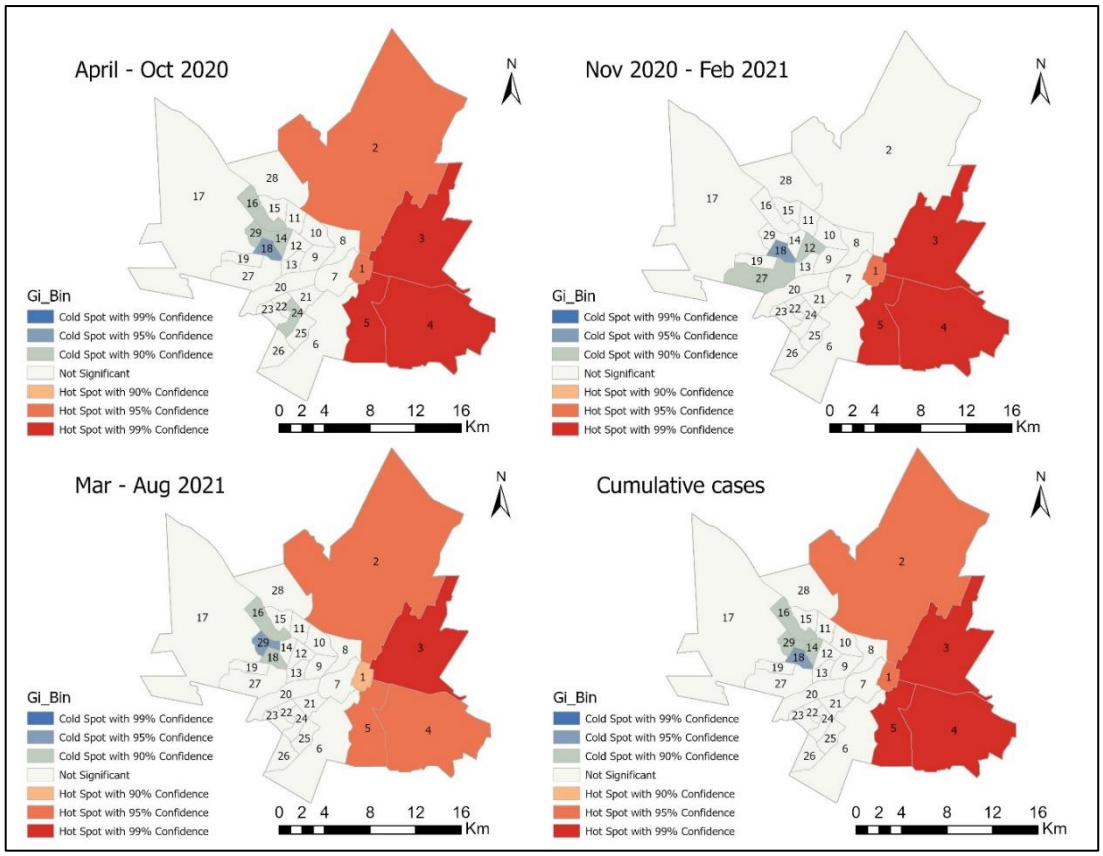


Figure 8: Hotspot analysis maps

In the second wave, a shift in the spatial distribution of the hot and cold spots was identified. Ward 2, which was initially a hotspot was no longer a hotspot in the second wave, while wards 1, 3, 4, and 5 remained significant hotspots. The cold spot moved southwards towards wards 12 and 27 while ward 18 remained the most significant. In the third wave, a movement of the cold spot towards the north was identified, although in the same area as in the first wave. The most significant cold spot in the third wave was found in ward 29 with 99% confidence. In the third wave, hotspots were identified in wards 1, 2, 3, 4, and 5 with ward 3 being the most significant hotspot at 99% confidence and ward 1 being the least significant hotspot with 90% confidence.

Using the cumulative COVID-19 cases data for the period April 2020 to August 2021, significant hotspots were identified in wards 1, 2, 3, 4, and 5. Wards 14, 16, and 29 are significant cold spots with 90% confidence, while ward 18 was the most significant cold spot with 95% confidence. Throughout the period, no clustering of the cold spot with a confidence of 99% was observed.

4.3 COVID-19 Spatiotemporal clustering and hotspots

4.3.1 Space-time cube

A 3D visualization of COVID-19 cases in Bulawayo using the input of a Space-time cube of all monthly COVID-19 cases recorded in Bulawayo wards for the April 2020 – August 2021 period is shown in Figure 9. In each bin of the wards cube, the count of COVID-19 cases was calculated for each month, and the trend for bin values across time at each location was measured using the Mann-Kendall statistic.

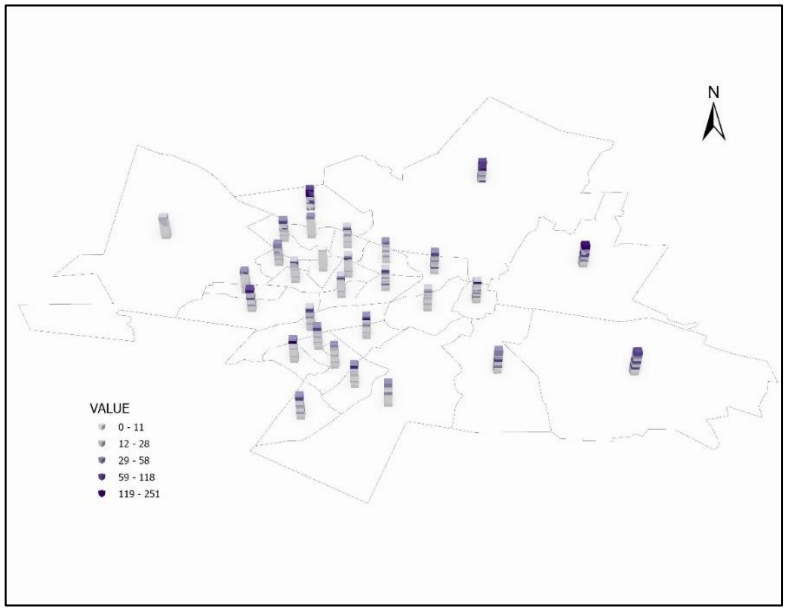


Figure 9: Space-time cube visualization

The Space-time cube visualization showed that majority of cases were found in the eastern wards of Bulawayo. From each ward bin time series, one can easily identify the first, second, and third wave of COVID-19 cases except for ward 17 which recorded a small number of cases.

4.3.2 Time series clustering analysis

Time series clustering analysis was conducted to group locations with similar temporal trends from April 2020 to August 2021. Based on the pseudo F-statistic table below, the optimal number of clusters selected was seven, which had the highest pseudo-F-statistic of 18.518.

Table 4: Pseudo F-statistic summary

Number of clusters	Highest Pseudo F
2	17.209
3	17.180
4	16.654
5	18.012
6	17.465
7	18.518
8	16.730
9	14.578
10	14.639

The chart below shows the average time series per cluster for the seven clusters. For the period April to October 2020, it was noted that the highest time series came from cluster 6 which consists of wards 4 and 5. Most COVID-19 positive cases in the second wave came from ward 3 which had the highest time series even in the third wave. The rest of the clusters had average to low time series with ward 17 having the lowest values in its time series throughout the whole period.

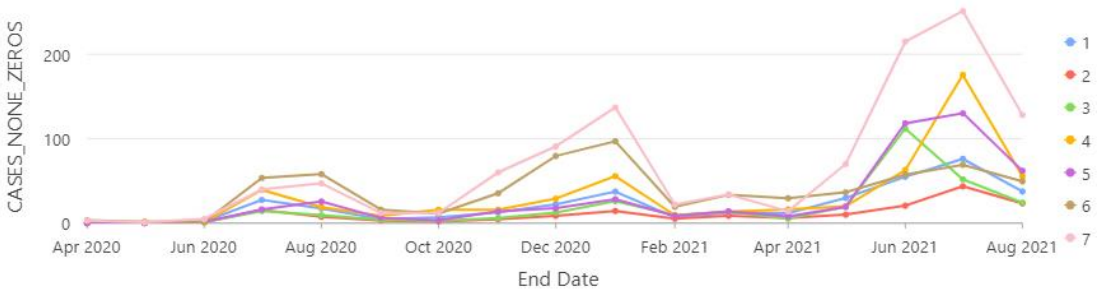


Figure 10: Average time series per cluster graph

Table 5 below shows the Mann-Kendall trend analyzes for the seven clusters. It was noted that all the time series clusters had an increasing trend and were all significant at alpha = 0.05. This result was like the overall monthly time series for all the COVID-19 cases in fig 5, which also represented an increasing trend. The three waves of COVID-19 cases were also visible from the clustered time series graph.

Table 5: Trend statistics for average time series (Mann-Kendall)

Cluster-ID	Direction	Statistic	p-value
1	Increasing	3.2982	0,0010
2	Increasing	3.4631	0,0005
3	Increasing	3.3807	0,0007
4	Increasing	3.0122	0,0026
5	Increasing	3.2158	0,0013
6	Increasing	2.2263	0,0260
7	Increasing	3.2982	0,0010

Most locations were classified under cluster 1 and cluster 2 with 7 and 8 wards, respectively, out of the 29. But, fewer of the wards were classified under clusters 6 and 7 (Table6).

Table 6: Number of locations per cluster

Cluster-ID	Number of locations
1	8
2	7
3	5
4	3
5	3
6	2
7	1

Figure 10 below shows the spatial distribution of the time series clusters. Most wards, which were under cluster 8, were found in the central and southern wards of Bulawayo. Wards classified under cluster 2 were located towards the western parts of Bulawayo. Clusters 6 and 7 were found in the eastern wards of Bulawayo.

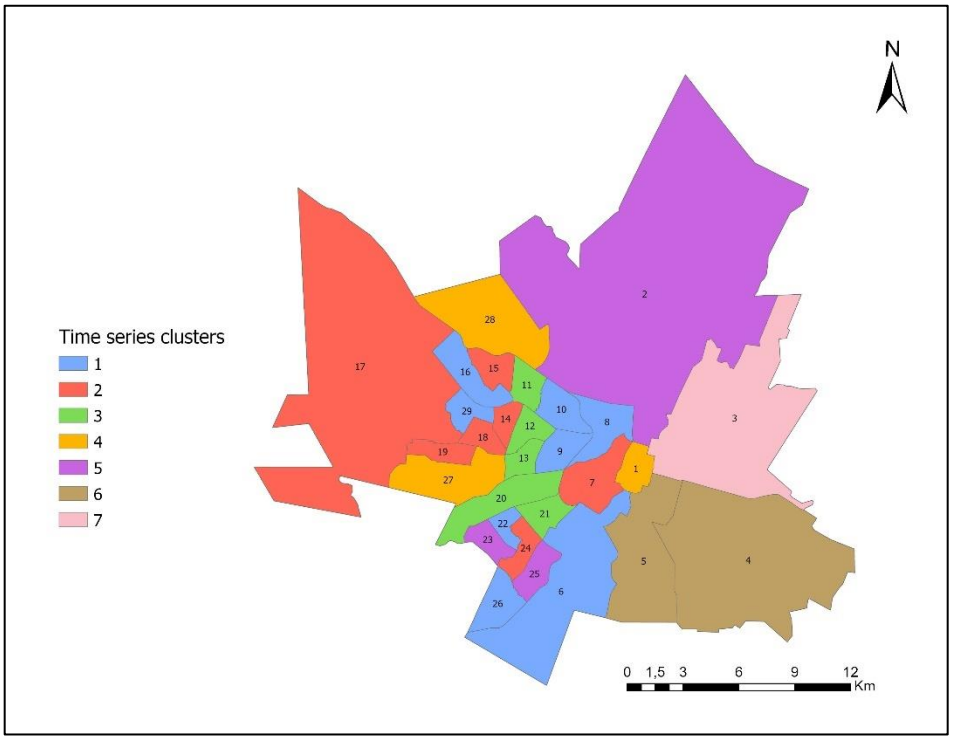


Figure 11: Time series clustering map

4.3.3 Local Outlier analysis

A local outlier analysis of the 29 locations based on the STC of positive COVID-19 cases was conducted and the analysis results are presented in Table 7 below. The spatiotemporal clustering pattern of COVID-19 was dominated mainly by a multiple clustering pattern, with 14 (48,28%) of the locations falling under that category.

Table 7: Number of locations per Local outlier analysis category

Category	Number of locations	% of locations
Never Significant	5	17,24
Only high-high cluster	1	3,45
Only low-high outlier	0	0,00
Only low-low cluster	9	31,03
Only high-low outlier	0	0,00
Multiple Types	14	48,28

Nine (31,03%) locations containing low-low clusters and one (3,45%) location being a high-high cluster was presented. Five of the wards (17,24%) in Bulawayo were insignificant. The results in Table

7 suggested that the outbreak presented a weak clustering pattern during the period of study, and the dominant spatial and temporal clustering pattern comprised multiple types.

A visual of the local outlier analysis is shown in Figure 12 . Most single clustering patterns and multiple types were distributed in the central wards of Bulawayo.

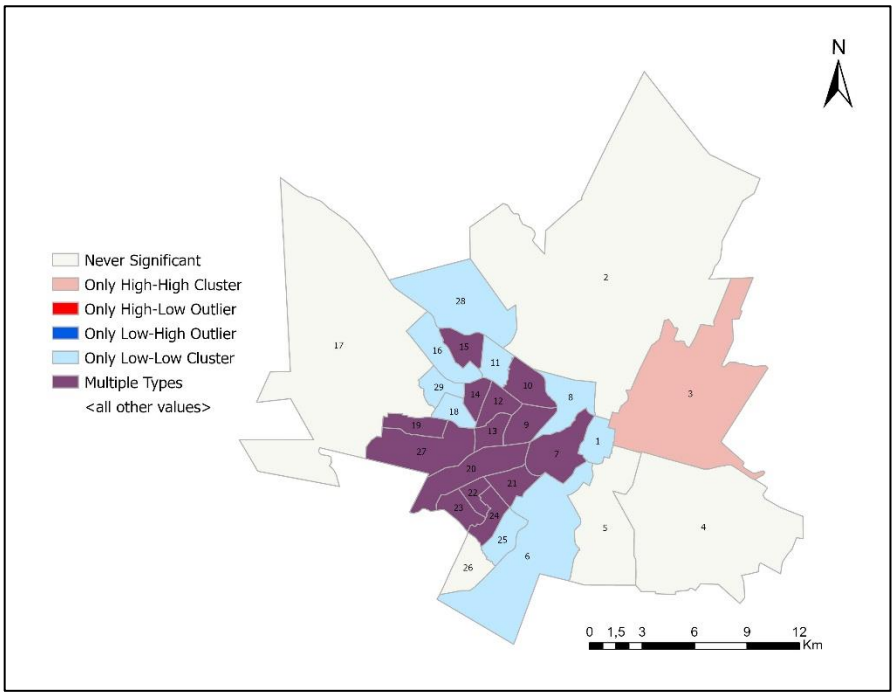


Figure 12: Local Outlier analysis map

The only high-high clustering pattern was found in the eastern side of Bulawayo in ward 3. Multiple type patterns were present in wards 12, 13, 14, 20, 21, 22, 23, and 24. Areas of low-low clustering were identified in the northern wards and southern wards in Bulawayo, which includes wards 11, 16, 28 in the northern parts and wards 1, 6, and 25 towards the south.

Many locations had multiple and significant only low-low clustering patterns, indicating that COVID-19 was still present at a low level throughout the region.

4.3.4 Emerging hotspot analysis

An Emerging hotspot analysis was run using the Space-time cube created for Bulawayo COVID-19 for the period April 2020 to August 2021. The results of the Emerging hotspot analysis categories are shown below in Table 8.

Table 8: Number of locations per Emerging hotspot analysis category

Patterns	Hot	Cold
New	0	0
Consecutive	13	0
Intensifying	0	0
Persistent	0	0
Diminishing	0	0
Sporadic	2	0
Oscillating	0	0
Historical	0	0

Fifteen out of 29 locations showed the spatiotemporal trend of hot spots and none of the locations showed trends in cold spots. Of the 15 locations which showed a trend, 13 were consecutive hotspots, a single uninterrupted run of hot time step intervals, comprised of less than 90% of all intervals. Two of the locations were identified as sporadic, with sometimes step intervals identified as hot. The rest of the areas did not show significant spatial and temporal trends.

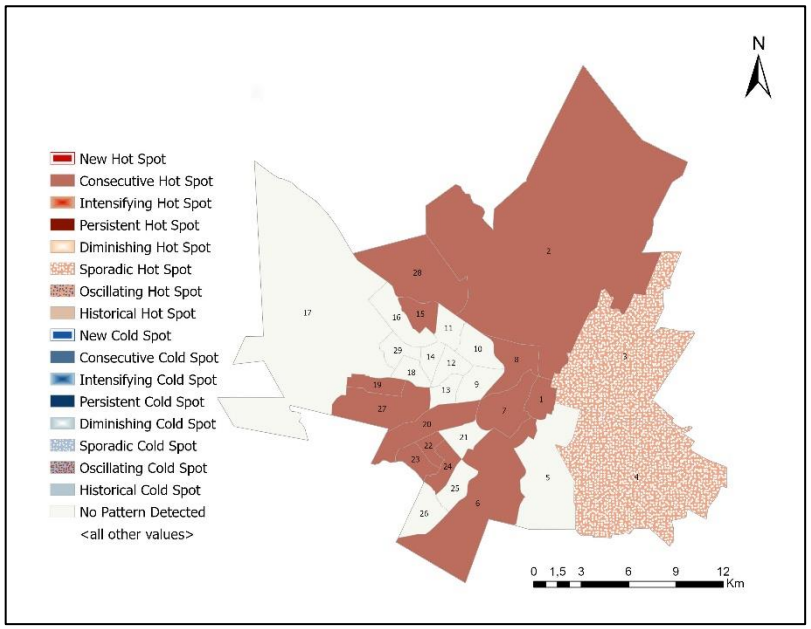


Figure 13: Emerging hotspot analysis map

Figure 13 shows the spatial distribution of the emerging hotspots. The sporadic hotspots were found to be concentrated in the south-eastern sides of Bulawayo. The consecutive hot spots were found mainly in the central parts of Bulawayo. No significant intensifying, new hot spots, persistent hotspots

were found, indicating that the COVID-19 situation in Bulawayo was stable and developing in a better direction.

4.4 Space-time forecasting of COVID-19 cases

4.4.1 Exponential Smoothing vs Forest-based forecasting

To effectively forecast COVID-19 cases for September 2021 in each ward, the Holt-Winters exponential smoothing approach was employed to dissect the time series at each ward of the space-time cube into seasonal and trend components. A Forest-based regression forecast was also done to predict COVID-19 cases for September and to compare the two prediction models. The forecasting models used one forecast time step and one timestep was excluded for validation.

Figure 14 below shows two maps of the forecasted COVID-19 cases for September 2021 using the two forecasting methods.

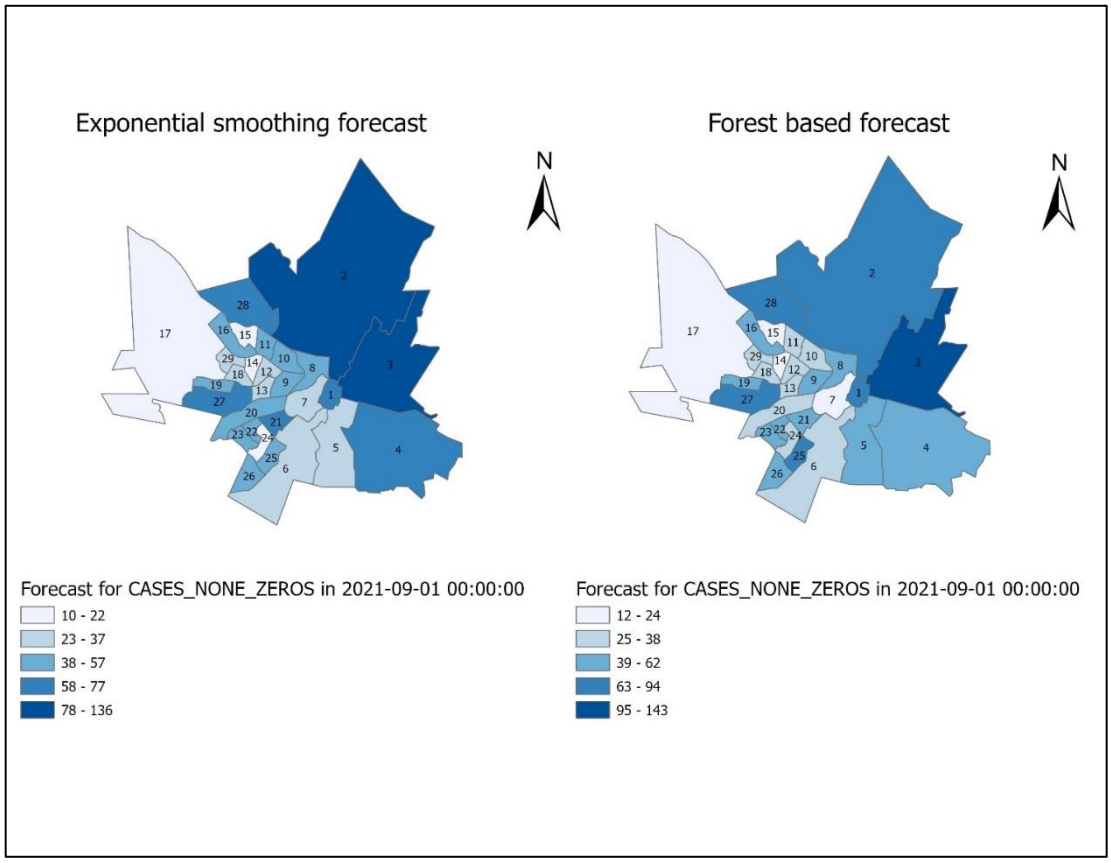


Figure 14: Exponential smoothing and Forest-based forecasting

The exponential smoothing map predicts that most cases came from wards 2 and 3 with between 78 to 136 cases in September 2021. Wards 14, 15, 17, and 24 had the least number of cases with less than

22. Most wards found in the central parts of Bulawayo were predicted to have between 23 to 77 COVID-19 cases each.

The Forest-based map predicted that the highest number of cases came from ward 3 with COVID-19 cases ranging between 95 to 143. Wards 7, 14, 15, and 17 were predicted to have the least number of cases of less than 24.

The forecast root mean square error (RMSE) was calculated to measure the fit of the exponential smoothing model and the Forest-based model to each time series. The Forecast RMSE evaluated how much the model's fitted COVID-19 cases varied from the raw time series COVID-19 caes. The Forecast RMSE for both models of the COVID-19 cases data in Bulawayo is shown in Table 9 below.

Table 9: Forecast RMSE for exponential smoothing vs Forest-based forecasting

Forecast cube	Category	Min	Max	Mean	Median	Std. Dev
Exponential smoothing	Forecast RMSE	5,72	48,89	21,17	20,33	10,36
Forest-based forecast	Forecast RMSE	3,14	30,46	11,86	11,03	5,93

The Forecast RMSE values for the Forest-based forecast were found to be relatively lower than those of the exponential smoothing forecasting. This means that Forest-based forecasting of COVID-19 in Bulawayo provided a better fit than exponential smoothing in fitting the values from the model to the raw time series values.

The Forecast RMSE solely assessed how well the exponential smoothing model fits the raw data. It did not assess how accurate the forecast model was at predicting future values. The validation model deals with this issue. The validation model was used to see how successfully each time series's forecast model could predict future values. The validation model was critical because it could directly compare expected values to raw values to assess how well the exponential smoothing model forecasted. The Validation RMSE for both models of the COVID-19 cases data in Bulawayo is shown in Table 10 below.

Table 10: Validation RMSE for exponential smoothing vs Forest-based forecasting

Forecast cube	Category	Min	Max	Mean	Median	Std. Dev
Exponential smoothing	Validation RMSE	0,03	114,00	23,60	13,53	27,37
Forest-based forecast	Validation RMSE	1,42	51,80	15,10	12,56	11,36

The Validation RMSE values for the Forest-based forecast were found to be relatively lower than those of the exponential smoothing forecasting. This meant that Forest-based forecasting of COVID-19 in Bulawayo provided a better forecast of future values of each time series than exponential smoothing.

4.4.2 Evaluating forecasts by location

A further analysis was done to select the most accurate between Exponential smoothing and Forest-based forecasting results for each ward of the COVID-19 space-time cube. This was done by identifying the forecast method that provided the smallest validation and forecast root mean square error (RMSE) and the output map is shown in Figure 15. With the same time series data, we employed two forecasting models and chose the best forecast for each site.

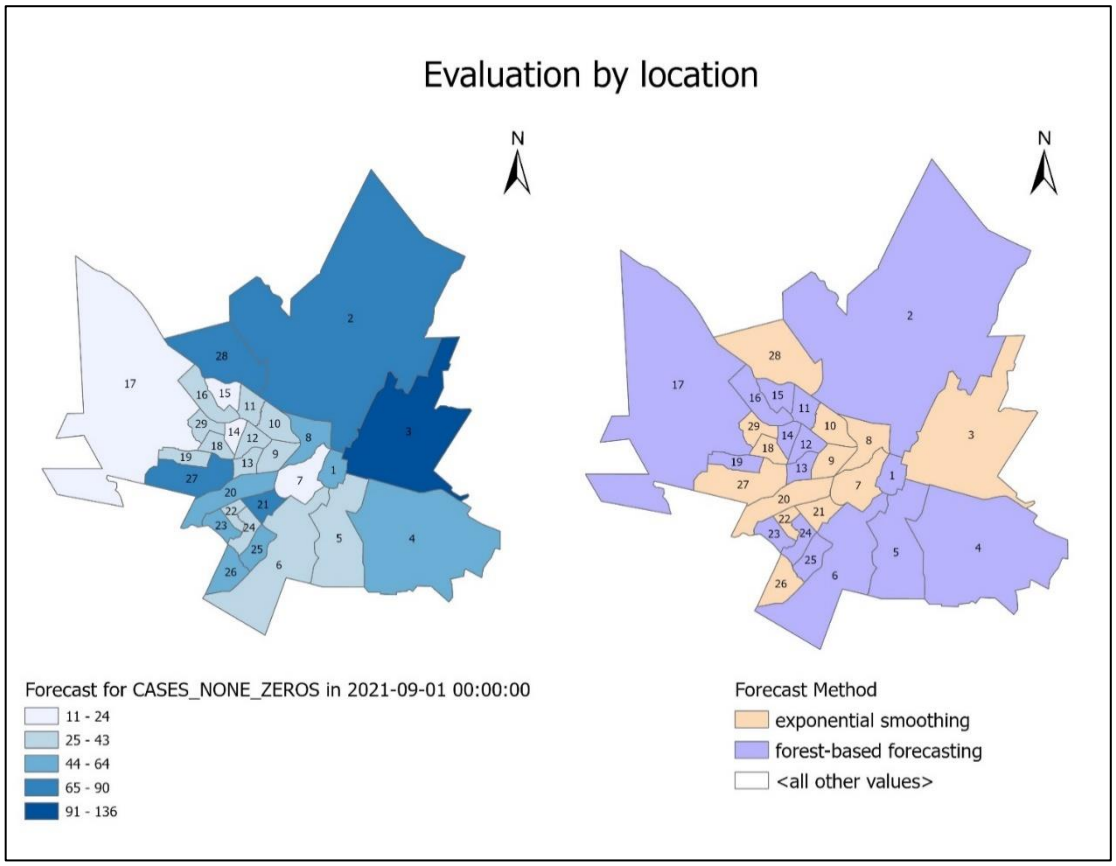


Figure 15: Combined forecasting showing evaluation by location

After evaluating the two forecasting methods, based on the validation RMSE, a combined output map of COVID-19 cases predicted for September 2021 was produced. The output map showed that the highest number of cases would come from ward 3 followed by wards 2, 21, 27, and 28. Most wards

found in the central area of Bulawayo had between 25 to 43 cases. Wards 7, 14, 15, and 17 had the least number of predicted cases.

Thirteen out of the 29 wards in Bulawayo were predicted using the exponential smoothing forecast and 16 of the wards using Forest-based forecasting.

5. Discussion

The research utilized geospatial techniques to map and examine the spatial and temporal patterns of COVID-19 in Bulawayo, Zimbabwe. From April 2020 to August 2021, the spatiotemporal distribution characteristics and shifting trends of cases were found. A time series analysis of COVID-19 reported cases was done to determine trends. The direction and size of the trend were then determined using the Mann Kendal trend test and the Sen slope. For the first three waves of the disease, a hotspot analysis, and a Cluster and Outlier analysis were used to conduct a 2D analysis of the spatial distribution and clustering of COVID-19 cases. A Space-Time Cube of COVID-19 cases was then created to enable a 3D analysis of COVID-19 in space and time. An Emerging hotspot analysis and a Local Outlier Analysis were used to identify the hotspot clusters temporally and determine the trends of the hotspots. COVID-19 occurrences were then forecasted spatially for September 2021 using Exponential smoothing and Forest-based forecasting. An evaluation by location was done to produce a combination of the two prediction models.

The temporal analysis shows there was a steadily increasing trend in COVID-19 cases in Bulawayo. The resurgence of COVID-19 variations could be one contributing factor. According to Murewanhema et al. (2021), COVID-19 cases in Zimbabwe were dominated by two highly contagious COVID-19 variants: the Beta version, which was initially found in South Africa, and the Delta variant, which was originally reported in India.

Another factor to consider was the rise in testing services. According to the WHO, Zimbabwe increased the use of antigen quick diagnostic tests to bolster its COVID-19 response in May 2021, reaching a daily rate of 4000 tests — a fourfold increase in just two months after the approach was debuted in November 2020. Because of the growth in testing, there is now a national coverage of testing and prompt reporting (WHO Africa, 2021).

The temporal analysis also revealed a two-wave pattern of reported cases in Bulawayo, where the first wave is identified around May to August and another from November to around January. The first wave of cases could be linked to winter as it is the cold and flu season when respiratory infections spread as people are more frequently inside and interacting nearby. According to Scoones (2021), the

surge of COVID-19 cases in July 2021 could be because of the relaxation in measures, including the day-to-day practices of hygiene that had occurred. People had gone back to a (nearly) normal life and abandoned wearing masks and had attended large gatherings of weddings, funerals, and church services. According to Murewanhema et al., (2020), the high increase in the daily number of COVID-19 incident cases in the second wave was due to an increase in the number of returning residents.

The spatial analysis revealed that most cases were recorded in the eastern sides of Bulawayo with ward 3 having the highest number of cases. Significant high-high clusters and hotspots were identified. But there were few cases recorded towards the central and western wards in Bulawayo. Low-low clusters and significant cold spots were identified in the western wards of Bulawayo. One explanation for this could be the high population density in the eastern suburbs compared to those in the west. Another explanation could be the high number of travelers reaching the area from outside the country including deportees. According to the City of Bulawayo (2020) in May 2020, 2 764 travelers came into Bulawayo to return to their homes. Deportees made up 31.6% of the 2 764 travelers that came through Bulawayo Polytechnic College and the United College of Education.

Due to the aggregation of each bin's positive case locations, 3D visualizations can help reveal hotspot/cold spot changing patterns in-depth. Another benefit is that policymakers can monitor and evaluate policies implemented at specific times in the bin about policies and actions taken to prevent the spread of COVID-19; however, the downside of the STC model's 3D visualization is that it can only see each cube from one side, making it difficult to understand and analyze hotspot patterns. From the emerging hotspot analysis, sporadic hotspots were discovered in the south-eastern parts of Bulawayo. Consecutive hot spots were also discovered in the central parts of Bulawayo.

The study proved the possibility to predict COVID-19 cases for future time steps using the exponential smoothing forecast and the Forest-based forecasting. The prediction of the spatial distribution of cases may assist the local authority in making effective and well-informed decisions to create and implement ward-targeted COVID-19 control policies in Bulawayo. There is, however, a need to incorporate other variables in the prediction models so that there is a better explanation to the occurrence of COVID-19. As seen from literature, on a worldwide scale, Wang et al. (2021) predicted the spatiotemporal distribution of COVID-19 using two climatic factors whilst Bariotakis et al. (2020) looked at 19 variables in explaining COVID-19. As COVID-19 progresses, more complicated and refined virus spread prediction models may be constructed.

6. Conclusion

The COVID-19 hotspots in Bulawayo were explored and located using geospatial methodologies and spatial statistics, setting the framework for further geospatial analysis, risk factor evaluation, epidemic monitoring, and management. This study's findings demonstrate the importance of geospatial mapping and spatial statistics, which can swiftly show and build maps by highlighting the geographic and temporal fluctuation of susceptible locations. The hotspot locations were identified, and recommendations were made for crucial preparedness, readiness, and response activities in the face of such a health crisis. Forecasts about where and when the disease will occur may be of great usefulness for decision-makers, as they give them the time to act on the local public health systems and better plan health policy interventions. Forecasting allows policymakers to be proactive and not reactive, which is critical for an effective response. This study will assist government agencies and communities in making effective and well-informed decisions to create and implement COVID-19 control policies in Bulawayo.

A major strength of the study is the use of both a 2D and 3D approach in trying to understand the spatial and temporal nature of COVID-19 incidence in Bulawayo. The approach helps in identifying hotspot/cold spots at different specified time intervals and looks at the overall temporal nature of the hotspots/cold spots.

A major obstacle faced during the research was the delay in receiving COVID-19 cases data. This could have been due to poor reporting systems from the facilities to the City of Bulawayo health department. The researcher had to wait for all data to be complete to start the analysis.

Due to the nature and sensitivity of the data, the study was done based on a ward level spatial resolution. Using station data at such a level may not produce the most accurate of results when conducting spatial analysis such as hotspot analysis and cluster outlier analyzes. The study also used a monthly temporal resolution which aggregates COVID-19 cases data monthly, leaving out some finer details in terms of occurrence.

In the spatiotemporal prediction analysis, only two forecasting techniques were used. More robust findings could have been obtained if more prediction models were applied. The research also offers short-term predictions based on the two prediction techniques which may be less effective in the long term. For such a study to be effective, the prediction findings should be put into practice early. This enables well-informed decisions in responding to COVID-19 and enforcing effective control measures in Bulawayo.

Notwithstanding these limitations, the study provides a meaningful backdated analysis of the spatial and temporal distribution of COVID-19 incidence from April 2020 to August 2021 at the ward level in Bulawayo. Thus, this study provides a baseline approach for epidemiologists to examine the causes of COVID-19 incidence for future research.

There is a need to apply such research in other cities in Zimbabwe, including Harare which has the highest number of COVID-19 cases in the country. Thus, this study provokes new research questions and provides the base for the up fourth comparative spatiotemporal research.

References

Anselin, L. (1995). Local indicators of spatial organization -LISA. *Research*, *27*(2), 1–25.

Arashi, M., Bekker, A., Salehi, M., Millard, S., Erasmus, B., Cronje, T., & Golpaygani, M. (2020). *Spatial analysis and prediction of COVID-19 spread in South Africa after lockdown*. *May*. http://arxiv.org/abs/2005.09596

Bansal, M. (2020). Cardiovascular disease and COVID-19. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(3), 247–250. https://doi.org/10.1016/J.DSX.2020.03.013

Bariotakis, M., Sourvinos, G., Castanas, E., & Pirintsos, S. A. (2020). Climatic influences on the worldwide spread of SARS-CoV-2. *MedRxiv*, *1*, 0–2.

Barman, A. (2020). *Time Series Analysis and Forecasting of COVID-19 Cases Using LSTM and ARIMA Models*. http://arxiv.org/abs/2006.13852

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chaora, B. (2020). *Impact of COVID 19 Lockdown on Micro, Small & Medium Scale Enterprises in Zimbabwe*. sivioinstitute.org

City of Bulawayo. (2020). *Mayor's Report on Covid-19 (21 May 2020)*. http://www.citybyo.co.zw/News/MayorsReportonCovid19_21May

Cordes, J., & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-Temporal Epidemiology*, *34*, 100355. https://doi.org/10.1016/j.sste.2020.100355

Djakaria, I., & Saleh, S. E. (2021). Covid-19 forecast using Holt-Winters exponential smoothing.

Journal of Physics: Conference Series, *1882*(1). https://doi.org/10.1088/1742-6596/1882/1/012033

Dutta, I., Basu, T., & Das, A. (2021). Spatial analysis of COVID-19 incidence and its determinants using spatial modeling: A study on India. *Environmental Challenges*, *4*. https://doi.org/10.1016/j.envc.2021.100096

Galasso, J., Cao, D. M., & Hochberg, R. (2022). A random forest model for forecasting regional COVID-19 cases utilizing reproduction number estimates and demographic data. *Chaos, Solitons & Fractals*, *156*, 111779. https://doi.org/10.1016/j.chaos.2021.111779

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/J.1538-4632.1992.TB00261.X

Gibson, L., Sibanda, B. N., Gwatec, O., Sibanda, S., Hargreaves, P., & Watmoughd, G. R. (2020). Bulawayo Operational Observatory and Mapping in response to COVID-19 social distancing requirements. *Proc. of SPIE Vol. 11527*, 115270A.

Hall, J., Jahic, A., Nayyar, S., Ortubia, A., Rivera, C., Tapia, H., Dardari, A. Al, Dhaliwal, M., García, O., Gilpin, R., Horvath, B., Intini, V., Lutz, B., López-calva, L. F., Roig, M., Saint-malo, I., & Slay, B. (2020). 2020 HUMAN DEVELOPMENT PERSPECTIVES COVID-19 AND HUMAN DEVELOPMENT : Assessing the Crisis , Envisioning the Recovery The team that created this report includes Pedro. *Undp*, 1–32. http://hdr.undp.org/sites/default/files/covid-19_and_human_development_0.pdf

Hirsch, R. M., & Slack, J. R. (1984). A Nonparametric Trend Test for Seasonal Data With Serial Dependence. *WATER RESOURCES RESEARCH*, *20*(6), 727–732. https://doi.org/10.1029/WR020i006p00727

Ho, T. K. (1995). Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, *1*, 278–282. https://doi.org/10.1109/ICDAR.1995.598994

Holt, C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10.

Hongfei, L., Calder, C. A., & Noel, C. (2007). Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.2007.00708.x

Kang, Y., Cho, N., & Son, S. (2018). Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation. *PLOS ONE*, *13*(5), e0196845. https://doi.org/10.1371/JOURNAL.PONE.0196845

Kendall, M. . (1975). *Rank Correlation Methods (4th edition)* (C. Griffin (ed.); 4th ed.).

Kim, S., & Castro, M. C. (2020). Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *International Journal of Infectious Diseases*, *98*, 328–333. https://doi.org/10.1016/j.ijid.2020.07.004

Kveladze, I., Kraak, M. J., & van Elzakker, C. P. J. M. (2018). Cartographic Design and the Space–Time Cube. *Https://Doi.Org/10.1080/00087041.2018.1495898*, *56*(1), 73–90. https://doi.org/10.1080/00087041.2018.1495898

Lak, A., Sharifi, A., Badr, S., Zali, A., Maher, A., Mostafavi, E., & Khalili, D. (2021). Spatio-temporal patterns of the COVID-19 pandemic, and place-based influential factors at the neighborhood scale in Tehran. *Sustainable Cities and Society*, *72*(May), 103034. https://doi.org/10.1016/j.scs.2021.103034

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, *13*(3), 245. https://doi.org/10.2307/1907187

Meals, D. W., Spooner, J., Dressing, S. A., & Harcum, J. B. (2011). *Statistical analysis for monotonic trends*. Tech Notes 6, November 2011. https://www.epa.gov/pollutedrunoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Ministry of Health and Child Care (MoHCC) Harare; Zimbabwe. (2021a). *Situation report COVID-19, Zimbabwe. 3 August 2021*.

Ministry of Health and Child Care (MoHCC) Harare; Zimbabwe. (2021b). *Situation report COVID-19, Zimbabwe. 4 August 2021*.

Mo, C., Tan, D., Mai, T., Bei, C., Qin, J., Pang, W., & Zhang, Z. (2020). An analysis of spatiotemporal pattern for COIVD-19 in China based on space-time cube. *Journal of Medical Virology*, *92*(9), 1587–1595. https://doi.org/10.1002/jmv.25834

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

Munanza, F., Arshad, S., Butt, I., & Arshad, S. (2021). Geospatial Clustering and Hot Spot Detection of COVID-19 Incidence in 2020 : A Global Analysis Geospatial Clustering and Hot Spot Detection of COVID-19 Incidence in 2020 : A. *International Journal of Geospatial and Environmental Research*, *8*(1).

Murewanhema, G., Burukai, T., Mazingi, D., Maunganidze, F., Mufunda, J., Munodawafa, D., & Pote, W. (2020). A descriptive study of the trends of COVID-19 in Zimbabwe from March-June

2020: policy and strategy implications. *The Pan African Medical Journal*, *37*(Suppl 1), 33. https://doi.org/10.11604/PAMJ.SUPP.2020.37.1.25835

Murewanhema, G., Dzinamarira, T., Chingombe, I., Mapingure, M. P., Mukwenha, S., Chitungo, I., Helena, H., Madziva, R., Ngwenya, S., & Musuka, G. (2022). Emerging SARS-CoV-2 Variants, Inequitable Vaccine Distribution, and Implications for COVID-19 Control in Sub-Saharan Africa. *COVID*, *2*(3), 341–349.

Ngwira, A., Kumwenda, F., Munthali, E. C. S., & Nkolokosa, D. (2021). Spatial temporal distribution of COVID-19 risk during the early phase of the pandemic in Malawi. *PeerJ*, *9*(April 2020). https://doi.org/10.7717/peerj.11003

OCHA. (2020). *ZIMBABWE Situation Report- 21 Aug 2020*. https://reports.unocha.org/en/country/zimbabwe/

Oxford Analytica. (2020). The WHO's COVID-19 pandemic declaration may be late. *Emerald Expert Briefings*.

Pourghasemi, H. R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S. R., Babaei, S., Khosravi, R., Etemadi, M., Ghanbarian, G., Farhadi, A., Safaeian, R., Heidari, Z., Tarazkar, M. H., Tiefenbacher, J. P., Azmi, A., & Sadeghian, F. (2020). Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *International Journal of Infectious Diseases*, *98*, 90–108. https://doi.org/10.1016/j.ijid.2020.06.058

Purwanto, P., Utaya, S., Handoyo, B., Bachri, S., Astuti, I. S., Sastro, K., Utomo, B., & Aldianto, Y. E. (2021). Spatiotemporal analysis of COVID-19 spread with emerging hotspot analysis and space-time cube models in East Java, Indonesia. *ISPRS International Journal of Geo-Information*, *10*(3). https://doi.org/10.3390/ijgi10030133

Rahman, M. R., Islam, A. H. M. H., & Islam, M. N. (2021). Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. *Modeling Earth Systems and Environment*, *7*(3), 2059–2087. https://doi.org/10.1007/s40808-020-00962-z

Ramírez-Aldana, R., Gomez-Verjan, J. C., Bello-Chavolla, O. Y., & Peterson, T. (2020). Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. *PLoS Neglected Tropical Diseases*, *14*(11), 1–15. https://doi.org/10.1371/journal.pntd.0008875

Scoones, I. (2021). *Zimbabwe faces a COVID-19 surge: what is happening in the rural areas? - Social Science in Humanitarian Action Platform*. https://www.socialscienceinaction.org/blogs-

and-news/zimbabwe-faces-a-covid-19-surge-what-is-happening-in-the-rural-areas/

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Shaharudin, S. M., Ismail, S., Samsudin, M. S., Azid, A., Tan, M. L., & Basri, M. A. A. (2021). Prediction of epidemic trends in covid-19 with mann-kendall and recurrent forecasting-singular spectrum analysis. *Sains Malaysiana*, *50*(4), 1131–1142. https://doi.org/10.17576/jsm-2021-5004-23

Sun, F., Matthews, S. A., Yang, T. C., & Hu, M. H. (2020). A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? *Annals of Epidemiology*, *52*, 54-59.e1. https://doi.org/10.1016/j.annepidem.2020.07.014

Sun, P., Lu, X., Xu, C., Sun, W., & Pan, B. (2020). Understanding of COVID-19 based on current evidence. *Journal of Medical Virology*, *92*(6), 548–551. https://doi.org/10.1002/jmv.25722

Wang, J., Tang, K., Feng, K., Lin, X., Lv, W., Chen, K., & Wang, F. (2021). Impact of temperature and relative humidity on the transmission of COVID-19: a modelling study in China and the United States. *BMJ Open*, *11*, 43863. https://doi.org/10.1136/bmjopen-2020-043863

Wang, Q., Dong, W., Yang, K., Ren, Z., Huang, D., Zhang, P., & Wang, J. (2021). Temporal and spatial analysis of COVID-19 transmission in China and its influencing factors. *International Journal of Infectious Diseases*, *105*, 675–685. https://doi.org/10.1016/j.ijid.2021.03.014

WHO Africa. (2021). *Rapid tests up Zimbabwe's COVID-19 diagnosis | WHO | Regional Office for Africa*. https://www.afro.who.int/news/rapid-tests-zimbabwes-covid-19-diagnosis

Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Http://Dx.Doi.Org/10.1287/Mnsc.6.3.324*, *6*(3), 324–342. https://doi.org/10.1287/MNSC.6.3.324

World Population Review. (2022). *Population of Cities in Zimbabwe*. https://worldpopulationreview.com/countries/cities/zimbabwe

Yang, W., Deng, M., Li, C., & Huang, J. (2020). Spatio-temporal patterns of the 2019-ncov epidemic at the county level in hubei province, china. *International Journal of Environmental Research and Public Health*, *17*(7), 1–11. https://doi.org/10.3390/ijerph17072563