Hindawi
International Journal of Mathematics and Mathematical Sciences
Volume 2024, Article ID 1591016, 17 pages
https://doi.org/10.1155/2024/1591016

Research Article

Modeling the Impact of Air Pollution and Meteorological Variables on COVID-19 Transmission in Western Cape, South Africa

John Kamwele Mutinda [1] **and Amos Kipkorir Langat** [2]

[1]*University of Science and Technology of China, Hefei, China*
[2]*Pan African University Institute for Basic Sciences, Technology and Innovation, JKUAT, Department of Mathematics, Nairobi, Kenya*

Correspondence should be addressed to Amos Kipkorir Langat; moskiplangat@gmail.com

Received 6 January 2024; Revised 17 February 2024; Accepted 8 March 2024; Published 24 April 2024

Academic Editor: Marco Costa

Understanding the factors that influence COVID-19 transmission is essential in assessing and mitigating the spread of the pandemic. This study focuses on modeling the impact of air pollution and meteorological parameters on the risk of COVID-19 transmission in Western Cape Province, South Africa. The data used in this study consist of air pollution parameters, meteorological variables, and COVID-19 incidence observed for 262 days from April 26, 2020, to January 12, 2021. Lagged data were prepared for modeling based on a 6-day incubation period for COVID-19 disease. Based on the overdispersion property of the incidence, negative binomial (NB) and generalised Poisson (GP) regression models were fitted. Stepwise regression was used to select the significant predictors in both models based on the Akaike information criterion (AIC). The residuals of both NB and GB regression models were autocorrelated. An autoregressive integrated moving average (ARIMA) model was fitted to the residuals of both models. ARIMA (7, 1, 5) was fitted to the residuals of the NB model while ARIMA (1, 1, 6) was fitted for the residuals of the GP model. NB + ARIMA (7, 1, 5) and GP + ARIMA (1, 1, 6) models were tested for performance using root mean square error (RSME). GP + ARIMA (1, 1, 6) was selected as the optimal model. The results from the optimal model suggest that minimum temperature, ambient relative humidity, ambient wind speed, $PM_{2.5}$, and NO_2 at various lags are positively associated with COVID-19 incidence while maximum relative humidity, minimum relative humidity, solar radiation, maximum temperature, NO, PM load, PM_{10}, SO_2, and NO_X at various lags have a negative association with COVID-19 incidence. Ambient wind direction and temperature showed a nonsignificant association with COVID-19 at all lags. This study suggests that meteorological and pollution parameters play a vital independent role in the transmission of the SARS-CoV-2 virus.

1. Introduction

1.1. Background Information. It has been more than two years since the first COVID-19 case was reported in late December 2019 in Wuhan, China [1]. On January 30, 2020, the International Committee on Taxonomy of Viruses (ICTV) recognized the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) as the responsible causative agent of the disease [2] and on March 11, 2020, the World Health Organisation (WHO) declared it an international public health emergency due to the rapid infections by the virus [3]. This disease is transmitted from one person to another through close contact droplets generated by an infected person when coughing, sneezing, or speaking in closed setups [4].

Dry cough, loss of smell, fever, and fatigue are some of the most common symptoms. Breathing problems, chest pain or pressure, shortness of breath, and loss of speech or movement are some of the more serious symptoms [5]. Patients with minor symptoms have mostly been treated effectively, whereas serious patients required intensive hospitalization and respiratory ventilation [6]. According to

the Johns Hopkins viral dashboard, as of April 12, 2022, there are 445,504,399 confirmed cases, 378,505,653 recoveries, and 6,015,762 fatalities globally [7].

In Africa, the reported first case of COVID-19 was on February 14, 2020, in Egypt, and on February 27, 2020, in Nigeria [8]. That was almost two months since the first reported case in China. In the early phase of the pandemic, most of the reported cases from African countries were through importation. While the pandemic was progressing, local transmission surpassed imported cases, and the doubling time shortened [9]. Currently, almost all new COVID-19 cases in Africa are from community transmissions, and there are 11,631,795 confirmed cases, 10,785,146 recoveries, and 251,749 deaths in Africa, as stated by the Johns Hopkins virus dashboard [7].

Previous studies have shown that meteorological and pollution parameters have affected the spread and thriving of several viruses [10]; for example, ambient temperature and relative humidity have shown an inverse association with the infection rate of influenza infection in Japan [11, 12] found that SAR-COV-2 virus has a seasonal oscillation of outbreak, suggesting a strong link between climatic conditions and virus transmission [13] showed that the spread of SAR-COV-2 is higher in winter than in summer.

Exposure to air pollution is considered the cause of several diseases and deaths around the globe [14]; therefore, it would be of great relevance to investigate its impact on COVID-19 transmission. Studies have proven for SARS-CoV-1 that air pollution can facilitate the spread of the virus and increase its persistence in the atmosphere [15]. In the United States, a study put in evidence that long-term exposure to a high concentration of particulate matter with an aerodynamic diameter of less than 2.5-micron ($PM_{2.5}$) increases the risk of mortality [16].

[17] sampled aerosols (air pollutants) from various locations to evaluate the aerodynamic characteristics of SARS-CoV-2. High concentrations of viral ribonucleic acid (RNA) were found in submicron aerosols, particularly in Wuhan Hospital Intensive Care Units (ICU) rooms, according to the findings. Long-term air-quality data was found to be significantly associated with incidences of COVID-19 in a study of 71 Italian provinces [5]. Another study found that long-term exposure to air pollution is associated with a variety of negative health outcomes, including greater fatality rates and hospital admissions [18].

Air pollution has been found to play an important role in the spread of infectious diseases. Poor air quality, for instance, has been attributed to the spread of severe acute respiratory syndrome diseases [19] and hence respiratory disorders such as asthma, chronic obstructive pulmonary disease (COPD), and lung cancer [20]. Despite the fact that there is substantial evidence to suggest pollution parameters play a crucial role in COVID-19 transmission, the number of data-driven studies investigating the association between air pollution parameters and COVID-19 transmission is still limited. It is therefore crucial to identify the key risk factors that influence COVID-19 transmission.

Zhang et al. [21] explored the correlation between meteorological factors and SARS-COV-2 transmission (2020). Precipitation, humidity, wind speed, and temperature all play a role in the propagation of the SARS-COV-2 virus, according to the findings. As a result, getting a better knowledge of the effects of meteorological factors on the propagation and survival of the SARS-COV-2 virus could be crucial in guiding the COVID-19 pandemic response.

COVID-19 severity in African countries has differed significantly from that in China, Europe, and other parts of the world due to diverse factors, spanning from demographical, epidemiologic, socioeconomic, and environmental implications [22, 23]. Since these parameters drive COVID-19's evolution, recent studies have focused on determining how each of these factors influences the virus's local transmission. Some studies have suggested that environmental factors, particularly meteorological and pollutant parameters, influence COVID-19 transmission [24].

The possible role of meteorological variables on COVID-19 transmission was investigated by Diouf et al. [25] in 16 nations in West and North Africa, including three climatic regions: the Sahel, Maghreb, and Gulf Guinea. Kendall nonlinear rank test and Spearman rank correlation test were utilized. The findings revealed a statistically significant negative association between COVID-19-confirmed cases and temperature in the Maghreb and Gulf of Guinea. Positive associations were discovered across the Sahel. Positive associations with specific humidity were reported over the Sahel and Gulf of Guinea, whereas negative associations were found over the Maghreb.

The role of meteorological variables and pollutants in the transmission of COVID-19 during the harmattan season in equatorial Africa was examined by Ogunjo et al. [26]. They studied the link between meteorological factors and air pollutants with COVID-19 incidence in seven Nigerian areas using Spearman and Pearson tests. COVID-19 incidence instances were shown to be highly associated with meteorological variables, according to the findings. In several provinces, temperature and humidity had a negative association with daily COVID-19 incidences. Their impact on COVID-19 transmission, however, was less than that of particulate matter. The COVID-19 incidence had a positive association with particulate matter, but a negative association with relative humidity.

In Dhaka, Bangladesh, Islam et al. [20] explored the correlation between COVID-19, air quality, and meteorological variables using the Spearman correlation test. The association between COVID-19 incidence and the covariates was also assessed using a generalised additive model (GAM) and a multiple linear regression (MLR) model. Particulate matter ($PM_2.5$), carbon dioxide (CO_2), and ozone (O_3) had a strong negative association with daily COVID-19 incidence cases. However, there was no significant association between COVID-19 incidences and nitrogen dioxide (NO_2). Some meteorological variables, on the other hand, were found to have a substantial association with COVID-19 incidences. Relative humidity was shown to have

a significant positive association, while atmospheric pressure was found to have a significant negative association. These results were consistent with the results of authors in [27–29].

Jiang et al. [30] focused on investigating the effect of ambient air pollutants and meteorological variables on COVID-19 incidence in four cities in China. The study integrated both multivariate Poisson regression and time series analysis to understand the correlation of the variables with COVID-19 cases. It was shown that the particulate matter ($PM_{2.5}$) and relative humidity were substantially associated with an increased risk of COVID-19 while particulate matter (PM_{10}) and temperature were substantially associated with a decrease in the risk of COVID-19. These results are similar to those of a study conducted by Liang et al. [31] on the association between human influenza cases and particulate matter ($PM_{2.5}$) concentrations.

Lolli et al. [5] conducted a study aimed to identify the impact of climate and air pollution on COVID-19 transmission in Italy. They used nonlinear Spearman and Kendall rank correlation tests to investigate how climatic and air pollution parameters were related to COVID-19 transmission in Milan and Florence, two important urban centers in Northern Italy. The study's major findings suggested that virus transmission is adversely associated with relative humidity and temperature.

In five Indian cities, the authors in [32] undertook an exploratory study to look into the association between meteorological factors and ambient air pollution, with SARS-COV-2 transmission and fatality rates. They used Spearman and Kendall rank correlation at 0.01 level of significance. The study's main findings showed that particulate matter was positively associated with COVID-19 incidence. It also backed up the theory that particulate matter above a certain threshold increases the likelihood of SARS-COV-2 transmission and mortality.

The impact of meteorological factors on the dynamics of the COVID-19 pandemic in Poland was investigated by the authors in [33]. The goal of this study was to investigate the association between COVID-19 dynamics and meteorological factors (relative humidity, temperature, sunshine duration, and wind speed) in Poland. The methods used were cross-correlation function, principal component analysis, and random forests. The results revealed that maximum temperature, relative humidity, sunshine duration, and mean daily temperature variability had a positive association with COVID-19 incidence.

In the case of the Western Cape, there is a dearth of research focusing on modeling the relationship between meteorological factors and pollution parameters with COVID-19 transmission; therefore, this study aims to address the uncertainty surrounding the transmission of COVID-19 by investigating the potential influence of ambient air pollutants and meteorological parameters, considering the unique context of Western Cape Province, South Africa, where high infection rates have been recorded [34]. The overarching objective is to develop a comprehensive model that encompasses both meteorological and air pollution factors and their impact on COVID-19 transmission in this region. Specifically, the study seeks to determine if COVID-19 cases in the Western Cape exhibit overdispersion and discern any significant monotonic trends, develop a predictive model to estimate COVID-19 incidence based on these variables, and utilize the model to analyze the association between COVID-19 incidence and meteorological variables and air pollutants in Western Cape. This research endeavor holds the potential to inform critical public health policies and provide insights that can be extrapolated to other regions with similar conditions and perhaps even extended to the study of other infectious diseases.

2. Data and Methods

2.1. Study Area. The Western Cape is a province in South Africa's southwestern region, bordering both the Indian and Atlantic oceans. Figure 1 below shows an overview of the Western Cape Province in South Africa. The Western Cape Province is located at a longitude of approximately 21.86°S and a latitude of about 33.23°S. It is the fourth-largest of South Africa's nine provinces, covering a total area of 129,449 km^2 and boasting a population of approximately 7 million people. This region is characterized by mountainous terrain and features a Mediterranean climate, characterized by hot, dry summers and moderate, wet winters, with minimal summer rainfall along the coastline. The average annual precipitation is around 380 mm, with some areas in the northwest receiving as little as 125 mm. The average temperature is around 23 °C, making it the coldest region in South Africa. The primary economic activities in the Western Cape include transportation and industrialization, as documented by Wikipedia [36]. As of April 12, 2022, there have been 673,698 confirmed cases of COVID-19 in the Western Cape, resulting in 21,903 deaths, 647,671 recoveries, and 15,213 reinfections, as reported by the source cited by the authors in [37].

2.2. Data. The data analyzed in this study comprised 19 variables among which there are ten meteorological variables and seven pollution variables. The COVID-19 data were accessed from the COVID-19 (2019-nCoV) Data Repository [38]. The data consist of daily case data, recovered and associated deaths from April 26, 2020, to January 12, 2021. The choice of the study period was influenced by data availability. The meteorological and air pollution data were retrieved from the South African Weather Service air quality information system [39]. Table 1 below gives a detailed description of the variables in the datasets.

2.3. Dealing with Missing Values. The data used in this study contained missing values for the SO_2 variable from July 10, 2020, to July 21, 2020. NO_2, NO, and NO_X variables contained missing values from September 23, 2020, to October 8, 2020. SO_2, NO_2, NO, and NO_X variables also recorded missing values from October 10, 2020, to November 26, 2020, with the largest number of missing values observed from December 31, 2020, to January 12, 2021, for all other variables except four out of eighteen variables. The random

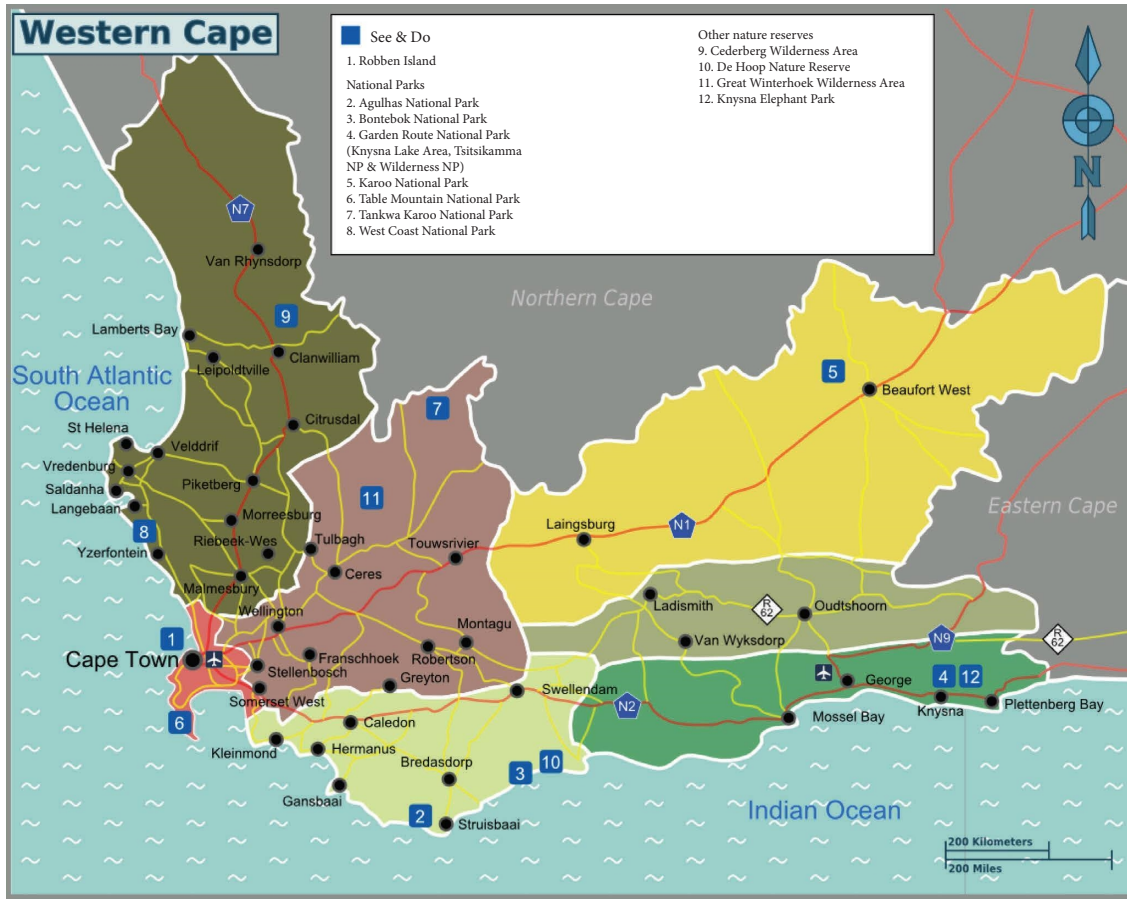

Figure 1: Map showing Western Cape Province [35].

Table 1: Description of variables in dataset.

Parameter	Description (units)	Temporal resolution
Incidence	COVID-19 cases (counts)	Daily
Deaths	COVID-19 deaths (counts)	Daily
TempMax	Minimum temperature (°C)	Daily
TempMIn	Maximum temperature (°C)	Daily
MaxHumidity	Maximum humidity (%)	Daily
MinHumidity	Minimum humidity (%)	Daily
Amb Wspeed	Normal wind speed (m/s)	Daily
Amb Wdirection	Direction of the wind (m/s)	Daily
Temperature	Normal temperature (°C)	Daily
Amb RelHum	Ambient relative humidity (%)	Daily
Solar radiation	Amount of incoming solar radiation (W/m^2)	Daily
Amb pressure	Ambient pressure (hPa)	Daily
SO_2	Concentration of sulphur dioxide ($\mu g/m^3$)	Daily
NO_2	Concentration of nitrogen dioxide ($\mu g/m^3$)	Daily
NO_X	Concentration of oxides of nitrogen ($\mu g/m^3$)	Daily
NO	Concentration of nitrogen monoxide ($\mu g/m^3$)	Daily
$PM_{2.5}$	Concentration of particulate matter of aerodynamic diameter of 5 μm ($\mu g/m^3$)	Daily
PM_{20}	Concentration of particulate matter of aerodynamic diameter of 10 μm ($\mu g/m^3$)	Daily
PM-load	Concentration of particulate matter ($\mu g/m^3$)	Daily

forest was utilized to impute the missing values of meteorological and pollution variables using multivariate imputation by chained equations (MICE). MICE uses a mean matching method to estimate missing values for continuous data and logistic regression to estimate missing values for binary data using random draws from independent normal distribution centered on means predicted from random forests. In the case of data used in this study, all variables were continuous. The imputation of missing values using R version 4.2.0.

2.4. Hypothesis Tests

2.4.1. Overdispersion Cameron and Trivedi (CT) Test. Cameron and Trivedi (1990) presented the CT test for detecting overdispersion in COVID-19 incidence, where H_0 is the equidispersion provided by $\mathbb{V}(Y|X) = \mathbb{E}(Y|X)$ based on the equation $\mathbb{V}(Y|X) = \mathbb{E}(Y|X) + \phi[\mathbb{E}(Y|X)]^2$ for negative binomial model and $\mathbb{V}(Y|X) = \mathbb{E}(Y|X)(1 + \phi\mathbb{E}(Y|X))^2$ for generalised Poisson model. A Poisson regression model should be estimated a priori in order to detect overdispersion in COVID-19 incidence at a specified level of significance. The Poisson regression model is defined as follows:

$$\begin{aligned} \log(\mathbb{E}[Y|X]) &= \log(\lambda) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n, \\ \mathbb{E}[Y|X] &= \lambda = e^{\beta' X}, \end{aligned} \tag{1}$$

where $\mathbb{E}[Y|X] = \lambda$ is the expected count, $\beta_0, \beta_1 \ldots \beta_n$ are parameters to be determined, and $X_1, X_2 \ldots X_n$ are the predictor variables. To perform the CT test, the following steps are used.

Step 1. Formulate the null and alternative hypotheses

$$\begin{aligned} H_0&: \phi = 0, \text{that is, there is equidispersion in the data,} \\ H_a&: \phi > 0, \text{that is, there is overdispersion in the data,} \end{aligned} \tag{2}$$

Step 2. Estimate the auxiliary OLS regression model without the intercept. The fitted values of λ for the first developed Poisson regression model are then used to compute the dependent variable Y^* as follows:

$$Y_i^* = \frac{[Y_i - \lambda_i]^2 - Y_i}{\lambda_i}, \tag{3}$$

where Y^* is the dependent variable, Y_i is the observed value, and λ_i is the fitted value. The auxiliary model in (3) sets λ as its single predictor variable is as follows:

$$Y_i^* = \beta \frac{\lambda_i}{\mathrm{H}}. \tag{4}$$

Step 3. The p value of the predictor variable λ_i is then examined using Student's t-test. When $p(|t|) > \alpha$, it is assumed that the data are equidispersed at a certain level of significance. In contrast, if $p(|t|) < \alpha$, then overdispersion is verified at a given level of significance α [40].

2.4.2. Ljung–Box Test. Ljung–Box test calculated the overall randomness based on the number of lags by examining the absence of autocorrelation up to specified lags. To perform the Ljung–Box test, the following procedure is used:

Step 1. Define the null and the alternative hypothesis as

$$\begin{aligned} H_0&: \text{the residuals of the model are random.} \\ H_a&: \text{the residuals of the model are autocorrelated.} \end{aligned} \tag{5}$$

Step 2. Compute the Ljung–Box test statistic defined by

$$Q_{LB} = n(n+2)\sum_{j=1}^{h} \frac{\rho^2(k)}{n-k}, \tag{6}$$

where n is the sample size, $\rho(k) = (\sum_{t=k+1}^{n}(y_t - \overline{y})(y_{t-k} - \overline{y})/\sum_{t=1}^{n}(y_t - \overline{y})^2)$ is the autocorrelation at time lag, k and h are number of lags, and Q_{LB} is the Ljung–Box test statistic.

Step 3. At a significance level α, the null hypothesis is rejected if

$$Q_{LB} > \chi^2_{1-\alpha,h}, \tag{7}$$

where χ^2 is the percentage point function of the chi-square distribution [41].

2.5. Modeling. Negative binomial regression and generalised Poisson regression models were used in this study. These models are used for count data that are overdispersed, that is, when the variance of the response variable exceeds the mean. The expected value of the response count is expressed as a linear combination of all other predictor variables.

2.5.1. Negative Binomial (NB) Regression Model. The negative binomial regression, also known as Poisson–Gamma mixture distribution, is defined by the probability function as

$$P[y_i|u_i, \theta] = \frac{\Gamma\left(y_i + \theta^{-1}\right)}{\Gamma(y_i + 1)\Gamma\left(\theta^{-1}\right)}\left(\frac{1}{1+\theta\mu_i}\right)^{\theta^{-1}}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i}, \tag{8}$$

where $\mu_i = t_i\mu$ and θ is the dispersion parameter. The parameter μ represents the average incidence rate of y per unit time of exposure t_i. The parameter μ is defined as the probability of a new occurrence of the event during a given exposure. The mean of y is determined by the exposure time t and a set of k regressors in a negative binomial regression model and is defined by

$$\log[\mathbb{E}(y_t)] = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_n X_{n,t} + \eta_t, \tag{9}$$

where $\log[\mathbb{E}(y_t)]$ is a linear function of n predictor variables $\mathrm{X}_{1,t}, \mathrm{X}_{2,t}, \ldots \mathrm{X}_{n,t}$. β_0 is the intercept and $\beta_1, \beta_2, \ldots . \beta_n$ are unknown parameters to be determined. The conditional mean function is the same as that of the Poisson distribution and it is

$$\mathbb{E}\left[y_t|x_{i,t}\right] = \mu_i = e^{\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_n X_{n,t} + \eta_t}, \tag{10}$$

and the variance $\mathbb{V}[y_t|x_{i,t}] = \mu_i(1 + \theta\mu_i)$ [42]. The dispersion factor is denoted by θ. The distribution is a Poisson distribution with parameter θ and is equidispersed when θ. When θ, the variance is smaller than the mean, and the count variable is underdispersed, and when θ, the count variable is overdispersed. In this study, the negative binomial (NB) model is of the following form:

$$\log[\mathbb{E}(y_t)] = \beta_0 + \sum_{k=1}^{6}\sum_{i=1}^{17}\beta_{(k,i)}X_{(i,t-k)} + \eta_t, \tag{11}$$

where $\mathbb{E}[y_t]$ is the expected COVID-19 incidence, β_0 is the intercept, $\beta_{(k,i)}$ is the coefficient of i^{th} variable at lag k, X_1, X_2, . . ., X_{17} are the pollution and meteorological variables, and η_t is the residual of the model. This model takes into consideration the lagged values of air pollution and meteorological variables. The average incubation period for COVID-19 is 6 days [43]. However, the incubation period for COVID-19 may be up to more than 6 days. The choice of the maximum length of incubation period in this study was taken to be 6 days. A stepwise model selection procedure based on the Akaike information criterion (AIC) was employed to drop models with the highest AIC values in the fitted negative binomial regression model [44]. This was performed by systematically skipping single variables in order to validate variables that have the least relevance and exclude them from further analysis.

2.5.2. Generalised Poisson (GP) Regression Model. A generalised Poisson (GP) regression model is a form of generalised linear model used to model count data which are random, underdispersed, overdispersed, or equidispersed [45]. A random variable y has a generalised Poisson distribution with parameters μ_i if it takes values $y = 0, 1, 2, \ldots$ with probability.

$$P[y_i|\mu_i,\phi] = \left(\frac{\mu_i}{1+\phi\mu_i}\right)^{y_i}\left(\frac{1+\phi y_i}{y_i!}\right)^{y_i-1} e^{-u_i\left(1+\phi y_i/1+\phi\mu_i\right)}. \tag{12}$$

The dispersion factor is denoted by ϕ. The distribution is a Poisson distribution with parameter μ_i and is equidispersed when $\phi = 0$. When $\phi < 0$, the variance is smaller than the mean, and the count variable is underdispersed, and when $\phi > 0$, the count variable is overdispersed. The mean and variance of the generalised Poisson distribution are calculated as $\mathbb{E}[y_i] = \mu_i$ and $\mathbb{V}[y_i] = \mu_i(1+\phi\mu_i)^2$, respectively. To develop the generalised Poisson regression model, first, the canonical link function is defined as follows:

$$\log[\mathbb{E}(y|X)] = \log(\mu). \tag{13}$$

The linear predictor is then defined as

$$\log[\mathbb{E}(y_t)] = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_n X_{n,t} + \eta_t. \tag{14}$$

So that

$$\log\left[\mathbb{E}\left(y_t|X_{i,t}\right)\right] = \log(\mu) = \log[\mathbb{E}(y_t)]. \tag{15}$$

Therefore, the expected count is expressed as

$$\mathbb{E}\left[\left(y_t|X_{i,t}\right)\right] = \mu = e^{\beta_0+\beta_1X_{1,t}+\beta_2X_{2,t}+\cdots+\beta_nX_{n,t}+\eta_t}, \tag{16}$$

where β_0 is the intercept, $X_{1,t}$, $X_{2,t}$, . . . $X_{n,t}$ are the values of the covariates at time t, β_1 , β_2, ... β_n are the coefficients of the predictor variables [46]. The generalised Poisson model fitted in this study is of the form

$$\log[\mathbb{E}(y_t)] = \beta_0 + \sum_{k=1}^{6}\sum_{i=1}^{17}\beta_{(k,i)}X_{(i,t-k)} + \eta_t. \tag{17}$$

where the variables and parameters are as in (11). One of the assumptions of NB and GP models is that the residuals of the fitted models must not be autocorrelated [47]. The Ljung–Box test is used to check the residuals for serial autocorrelation in the models. If they show autocorrelation, an autoregressive integrated moving average model (ARIMA) is fitted to the residuals of that model to capture the unexplained patterns that exhibited in the residuals. The optimum ARIMA model for the residuals is chosen based on the AIC criteria.

2.5.3. Autoregressive Integrated Moving Average (ARIMA (p, d, q)) for the Residuals. After the Ljung–Box test of the residuals of the models, ARIMA (p, d, q) models for the residuals were fitted. The ARIMA (p, d, q) model is a time series approach that takes a combination of autoregression, integration, and moving average. Autoregression shows that the time series is regressed with its lagged values as follows:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t, \tag{18}$$

where α is the intercept term, ϕ_1, ϕ_2,. . ., ϕ_p are parameters to be determined, and ϵ_t is a white noise $\sim \mathcal{N}(0, \sigma^2)$.

Integrated $I(d)$ means differencing taken at d times until the original series becomes stationary. A stationary time series has properties that are independent of the time at which it was observed. The first order of difference is provided by

$$\Delta y_t = y_t - y_{t-1} = (1-\beta)y_t, \tag{19}$$

and general form of difference order is

$$\Delta^d y_t = (1-\beta)^d y_t, \tag{20}$$

The moving average (q) takes the present value as a linear combination of all lagged forecast errors, where the error terms are the errors of the autoregression models of the respective lags. Equation (21) shows the MA model

$$y_t = \alpha + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}, \tag{21}$$

where α is a constant and θ_1, θ_2,. . . θ_q are parameters to be determined. The complete ARIMA (p, d, q) model is given by

$$\begin{aligned}\Delta^d y_t = {} & \phi_1\Delta^d y_{t-1} + \phi_2\Delta^d y_{t-2} + \cdots + \phi_p\Delta^d y_{t-p} \\ & + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}.\end{aligned} \tag{22}$$

The optimal values for p and q are those from which the AIC of the corresponding ARIMA (p, d, q) model is the least. The parameters of ARIMA (p, d, q) models are obtained by maximum likelihood estimation [47].

2.6. Model Selection

2.6.1. Using AIC Criterion. In the fitted negative binomial and generalised Poisson models, a stepwise model selection approach based on the Akaike information criterion was used to exclude models with the highest AIC values. This was achieved by systematically skipping single variables in order to validate variables with the least importance and eliminate them from further analysis. From the stepwise regression of the NB, GP, and residual ARIMA (p, d, q) models, the AIC was used to choose the best-fitted model. The step regression and residual models with the lowest AIC values were chosen as the most appropriate optimizing models from the available models. The AIC is defined as follows:

$$\text{AIC} = -2\ln(L) + 2k, \tag{23}$$

where L represents the maximum value of the log-likelihood function and k represents the number of independent variables [48].

2.6.2. Root Mean Square Error (RSME). The RSME is used to examine the model performance by determining how far the fitted values fall from the observed values using the Euclidean distance. The RSME was used to determine which of the models NB + ARIMA (p, d, q) and GP + ARIMA (p, d, q) was optimal. The RMSE value is defined by

$$\text{RSME} = \sqrt{\frac{\sum_{i=1}^{N}(y(i) - \widehat{y}(i))^2}{N}}, \tag{24}$$

where $y(i)$ is the observed i^{th} value, $\widehat{y}(i)$ is the fitted i^{th} value, and N is the number of observations [49].

3. Results and Discussion

3.1. Distribution of the Response Variable. Figure 2 illustrates the distribution of COVID-19 incidence during the study period. It can be observed from Table 2 that the response variable is skewed to the right with a mean of 932 and a variance of 894,840. With the variance being 960 times higher than the mean, the incidence count variable is overdispersed. The overdispersion was further confirmed by the overdispersion test at 5% level of significance (Lambda t-test score = 10.882, p value ≤ 0.001). Based on this property of the response variable, negative binomial and generalised Poisson models were considered in this study.

3.2. Results from NB and GP Models

3.2.1. NB Model. Stepwise regression was performed on the NB model in (11) to drop the nonsignificant predictors. The RSME, which measures the deviation of the observed and fitted counts, was 470.86 for the full model and 456.585 for the reduced model. Furthermore, the AIC was 3695.919 for the reduced model while it was 3770.92 for the full NB model. As a result, the reduced model was preferred for modeling the COVID-19 incidence in Western Cape. The NB regression model fits and the observed COVID-19 incidence are shown in Figure 3. The results suggest that the NB model failed to capture well the behavior of COVID-19 incidence during the first days of the study period.

Autocorrelation and partial autocorrelation functions of residuals of the fitted NB regression model are shown in Figure 4. It can be observed that the residuals are not a white noise and some patterns still existed in the remaining series. The results were further confirmed by the Ljung–Box test at the 5% level of significance (p value ≤ 0.001, Ljung–Box test statistic = 283.03). To capture the unexplained patterns exhibited in the residuals, ARIMA (p, d, q) model was fitted on residuals. The augmented Dickey–Fuller test (ADF) confirmed that there was a nonstationarity in the residuals of the negative binomial regression model (Dickey–Fuller test statistic = −3.3922, p value = 0.05627). The time series of the residuals was stationarised through first differencing. The ACF and PACF of differenced residuals are shown in Figure 5.

The possible parameters for autoregressive (AR) and moving average (MA) components of the ARIMA model were identified from and by changing various combinations of AR and MA components. Several ARIMA models for the residuals were tested. Based on the AIC, the ARIMA (7, 1, 5) was selected as the optimal model.

Diagnostic analysis of residuals of ARIMA (7, 1, 5) was done. The residuals were normally distributed, random, and had a constant mean of 0. All the assumptions of residuals were met by the optimal model as shown in Figure 6. The fitted values for ARIMA (7, 1, 5) were then added to the fitted values of the NB regression model. A plot of the fitted values of NB + ARIMA (7, 1, 5) and the observed counts of the COVID-19 incidence is shown in Figure 7. The fitted series from the NB and NB + ARIMA (7, 1, 5) model can capture well most of the patterns in the actual series. However, both models failed to properly capture the behavior of COVID-19 incidence during the first days of the study period.

GP model stepwise regression was performed on the GP model in (17) to drop the nonsignificant predictors. The root means the square error was 317.68 for the full model and the root mean square error of the reduced model is 311.1374. Furthermore, the AIC of the full NB model was 3732.411 while it was 3655.563 for the reduced model. As a result, the reduced model was preferred for modeling the COVID-19 incidence in Western Cape. The GP regression model fits the observed COVID-19 incidence well, and this is shown in Figure 8. It can be observed that the model fits approximately through the mean of COVID-19 incidence. GP regression model was able to capture well the behavior of COVID-19 incidence during the study period.

Autocorrelation and partial autocorrelation functions of residuals of the fitted GP model are shown in Figure 9. It can be observed that the residuals are autocorrelated and there were some patterns still existing in the remaining series. The results were further confirmed by the Ljung–Box test at 5% level of significance (p value ≤ 0.001, Ljung–Box test statistic = 148.79). To capture the unexplained patterns exhibited in the residuals, an ARIMA (p, d, q) model was fitted on residuals. The augmented Dickey–Fuller test (ADF)

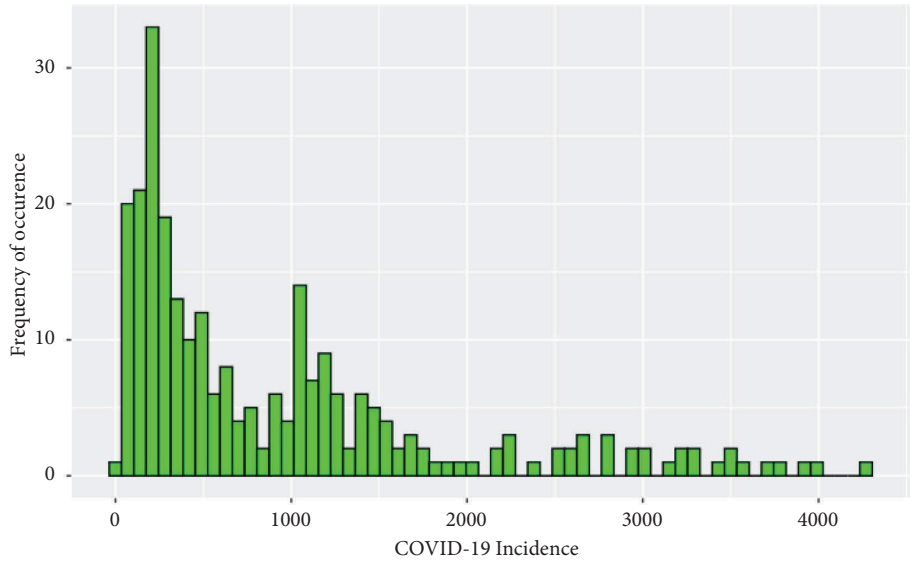

Figure 2: Distribution of the response variable.

Table 2: Descriptive statistics of COVID-19 incidence (response variable), air pollutants, and meteorological parameters (explanatory variables) from April 26, 2020, to January 12, 2021.

	Mean	Std	Min	25%	50%	75%	Max
Incidence	932.14	945.96	19.00	230.25	562.00	1253.50	4241.00
Deaths	32.29	38.90	0.00	7.00	21.00	40.00	265.00
TempMx	19.79	2.74	13.20	17.92	19.45	21.68	28.60
TempMn	13.87	2.95	6.50	11.90	13.60	16.18	21.30
MaxHumidity	87.65	5.92	69.00	85.00	90.00	92.00	97.00
MinHumidity	59.98	12.37	19.00	52.00	60.00	70.00	86.00
SO_2	−0.66	2.78	−7.03	−1.82	−0.21	1.09	6.21
NO_2	4.61	13.36	−27.56	−6.22	7.32	13.36	33.40
NO	2.70	3.29	−1.23	0.66	1.61	3.70	18.16
NO_X	7.31	14.89	−26.76	−4.63	8.59	17.07	48.28
$PM_{2.5}$	10.35	33.85	1.61	4.63	6.37	9.34	494.98
PM_{10}	24.11	95.09	4.32	11.44	14.92	18.68	1390.80
PM_Load	34.48	13.94	14.31	20.80	33.34	48.04	56.59
AmbWspeed	3.43	1.28	0.53	2.49	3.32	4.10	7.87
AmbWdirection	156.30	100.48	1.40	47.82	174.31	217.47	359.37
Temperature	16.80	3.05	9.00	15.00	16.00	19.00	25.00
AmbRelHum	86.82	10.16	55.48	80.84	89.21	94.64	100.87
SolarRadiation	211.14	99.97	25.54	135.55	185.64	284.23	470.07
AmbPressure	1016.8	5.19	1002.24	1013.53	1016.78	1020.06	1032.64

confirmed the residuals of the generalised Poisson regression model (Dickey–Fuller test statistic = −4.3922, p value = 0.06627) are nonstationary and it was overcome by first differencing. The time series of the residuals was stationarised through first differencing. The ACF and PACF of differenced residuals are shown in Figure 10.

The possible parameters for autoregressive (AR) and moving average (MA) components of the ARIMA (p, d, q) model were identified and by changing various combinations of AR and MA components, several possible ARIMA (p, d, q) models were tested. Several ARIMA (p, d, q) models for the residuals were tested. Based on the AIC, the ARIMA (1, 1, 6) was selected as the optimal model.

Then, the selected best model ARIMA (1, 1, 6) was checked for the validity of the assumptions. A diagnostic analysis of residuals was carried out. All the assumptions of residuals were met as shown in Figure 11, whereas the Ljung–Box test confirmed the nonautocorrelations at earlier lags at 5% significance level (p value = 0.8985, Ljung–Box test statistic = 4.881). The fitted values for ARIMA (1, 1, 6) were then added to the fitted values of the GP regression model. A plot of the fitted values of NB + ARIMA (1, 1, 6) and the observed incidence of COVID-19 incidence is shown in Figure 12. The fitted series from the GP and GP + ARIMA (1, 1, 6) models can capture uniformly most of the patterns in the actual series of COVID-19 incidence.

The RSME was used to identify the optimal model between NB + ARIMA (7, 1, 5) and GP + ARIMA (1, 1, 6). The RSME of the NB + ARIMA (7, 1, 5) is 456.9698 while that of GP + ARIMA (1, 1, 6) is 447.121 as a result GP + ARIMA

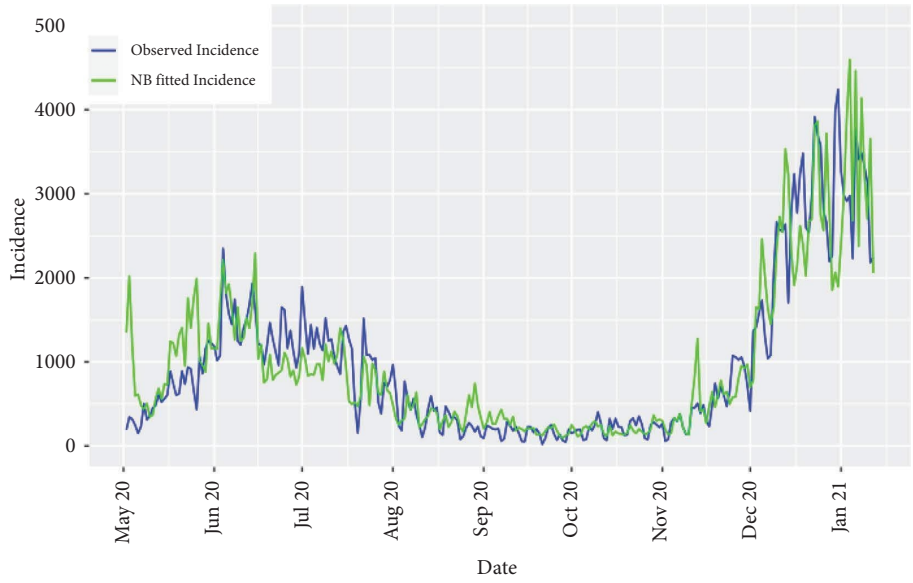

Figure 3: Observed and fitted values of COVID-19 incidence by NB regression.

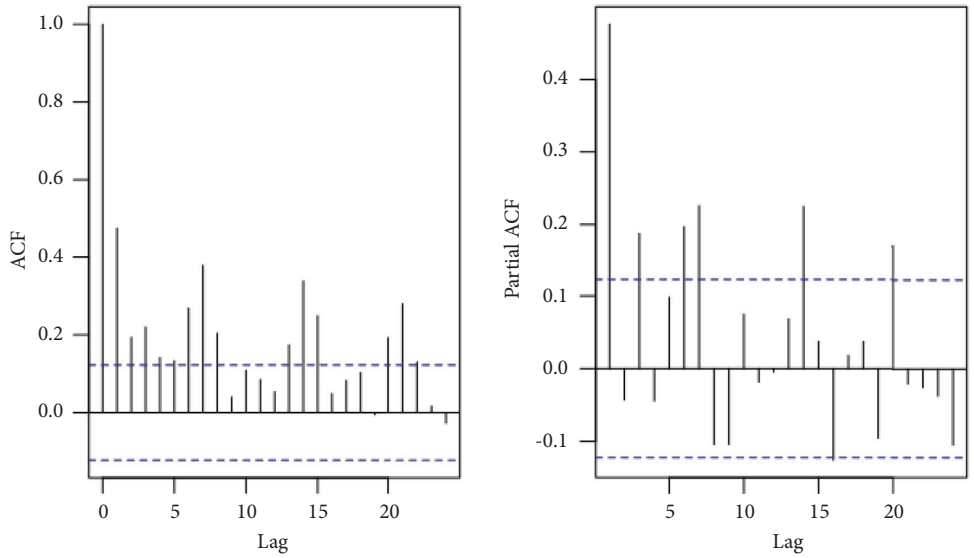

Figure 4: ACF and PACF plots of the residuals of the NB regression model.

(1, 1, 6) was selected as the optimal model with the minimum RMSE. Figure 13 shows a plot for the comparison of observed incidence counts, fitted values of NB + ARIMA (7, 1, 5) and fitted values of GP + ARIMA (1, 1, 6). By observing this plot, it can be seen that both models fit the incidence well. However, the GP + ARIMA (1, 1, 6) seems to be consistent with the observed incidence in Western Cape Province.

3.3. Discussion. The impact of pollution and meteorological variables on COVID-19 transmission was examined using NB and GP models. Overall, the GP + ARIMA (1, 1, 6) suggested that daily COVID-19 incidence was positively associated with minimum temperature, $PM_{2.5}$, NO_2, ambient relative humidity, and ambient wind speed at various lags. A negative association was observed between COVID-19 incidence and maximum humidity, minimum humidity, maximum temperature, PM_{10}, PM load, solar radiation, NO_X, and SO_2 at various lags as shown in Table 3. Care must be taken while interpreting the impact of air pollution and meteorological variables on COVID-19 incidence. With regards to a positive association, a positive coefficient indicates that an increase of that variable at lag k corresponded to an increase in the COVID-19 incidence holding all other variables constant, while for a negative association, a negative coefficient indicates that a decrease of the variable at lag k corresponds to a decrease of COVID-19 incidence holding all other variables constant.

The findings have shown that $PM_{2.5}$ at lag 1, lag 3, lag 4, and lag 6 suggested a positive association with incidence while PM_{10} and PM-load suggested a negative association with an incidence at lag 1, lag 3, lag 4. and lag 6, respectively. These findings

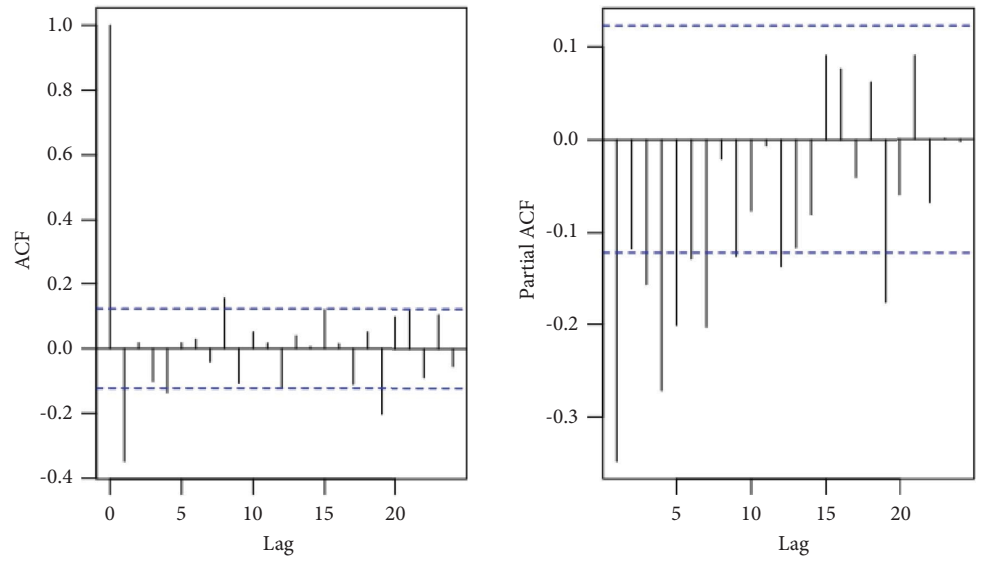

Figure 5: ACF and PACF plots of the differenced residuals of NB regression.

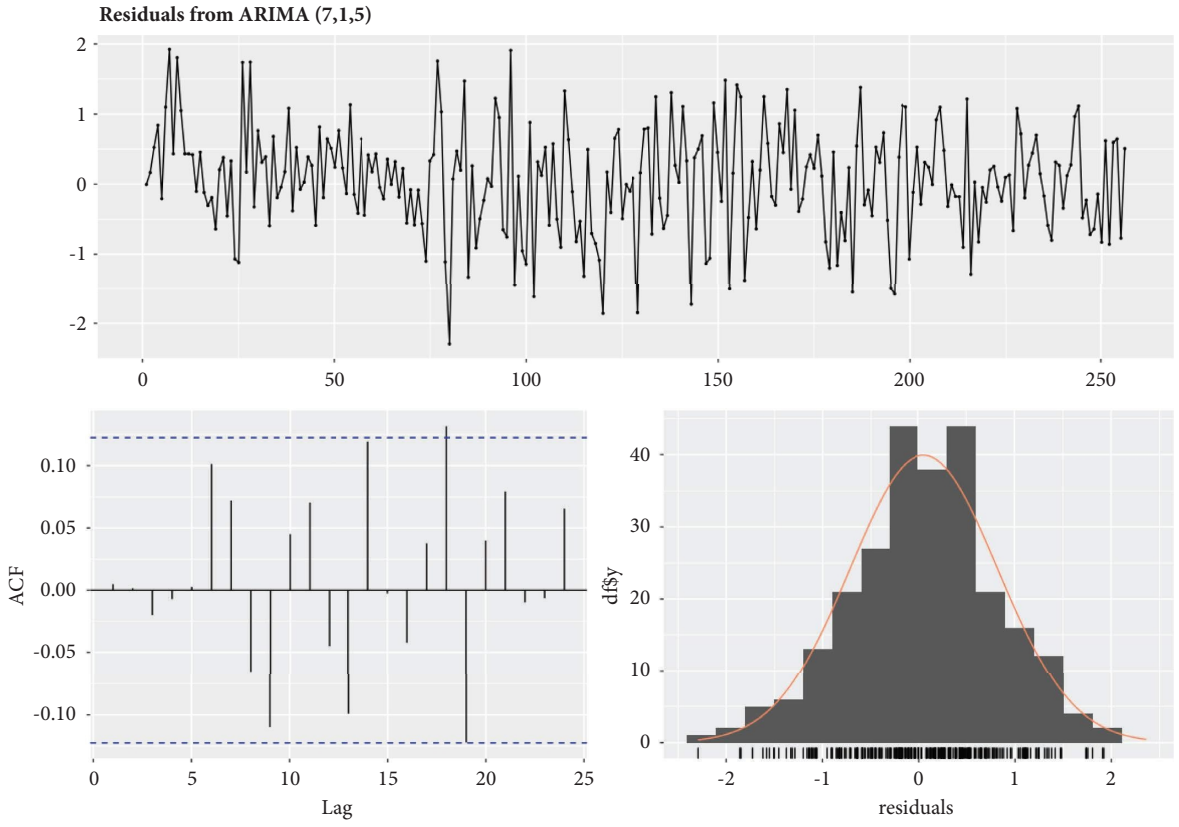

Figure 6: Residual analysis of ARIMA (7, 1, 5).

contradicted the findings of the authors in [50] which showed $PM_{2.5}$ is negatively associated with COVID-19 incidence and those of authors in [51] which showed that PM_{10} and PM-load were positively associated with COVID-19 incidence. Maximum temperature at lag 3 and lag 4 showed a negative association with incidence while the minimum temperature at lag 2, lag 3, and lag 6 showed a positive association with incidence. These findings are not in line with the findings of authors in [33, 52] which showed that maximum temperature and minimum temperature were positively associated with COVID-19 incidence. Average temperature showed no association with COVID-19 incidence. This was not in line with the findings of authors in [53] which suggested that average temperature was positively associated with COVID-19 incidence.

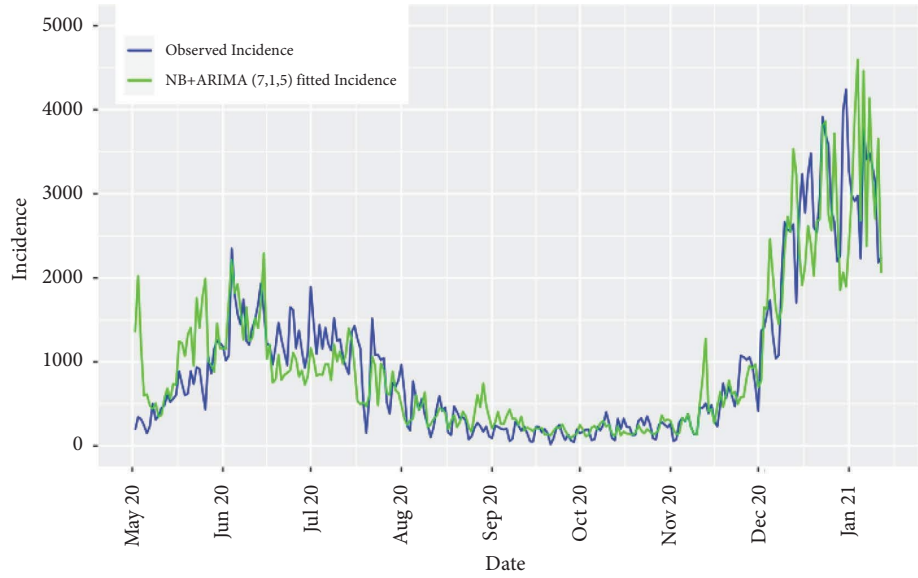

Figure 7: Actual and fitted values of negative binomial regression + ARIMA model (7, 1, 5).

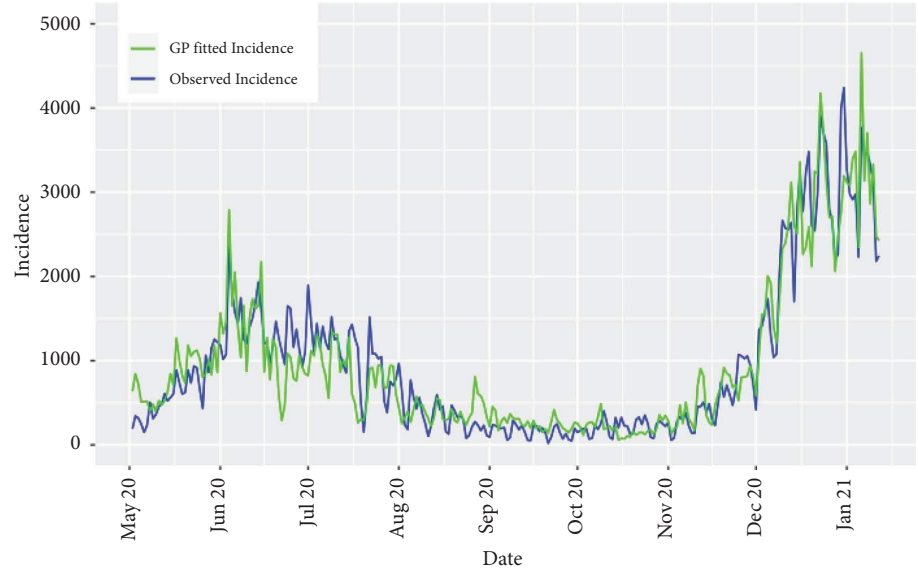

Figure 8: Actual and fitted values of COVID-19 incidence using the GP model.

Moreover, maximum relative humidity at lags 1, 2, and 5 and minimum relative humidity at lags 2, 3, and 6 exhibited a negative association with COVID-19 incidence, while ambient relative humidity at lags 2 and 5 showed a positive association with COVID-19 incidence. These results are consistent with those of authors in [54], which reported that maximum relative humidity was negatively associated with COVID-19 incidence, and authors in [20], which showed that ambient relative humidity was positively associated with COVID-19 incidence. With regards to NO, NO_2, NO_X, and SO_2, this study showed that NO suggested a negative association with incidence at lag 4 and NO_2 suggested a positive association with COVID-19 incidence at lag 5 while, NO_X and SO_2 showed a negative association with COVID-19 incidence at lag 5, lag 1 and lag 6, respectively. These findings were in line with the findings of authors in [24] which showed that NO_2 was positively associated with COVID-19 incidence and in contradiction with SO_2 which showed a positive association with COVID-19 incidence. In addition, ambient wind speed showed a positive association with COVID-19 incidence. This is consistent with the results of authors in [24] which showed that wind speed was positively associated with COVID-19 incidence. Solar radiation suggested a negative association with COVID-19 incidence at lag 4 and lag 6. This is in line with the findings of authors in [55, 56].

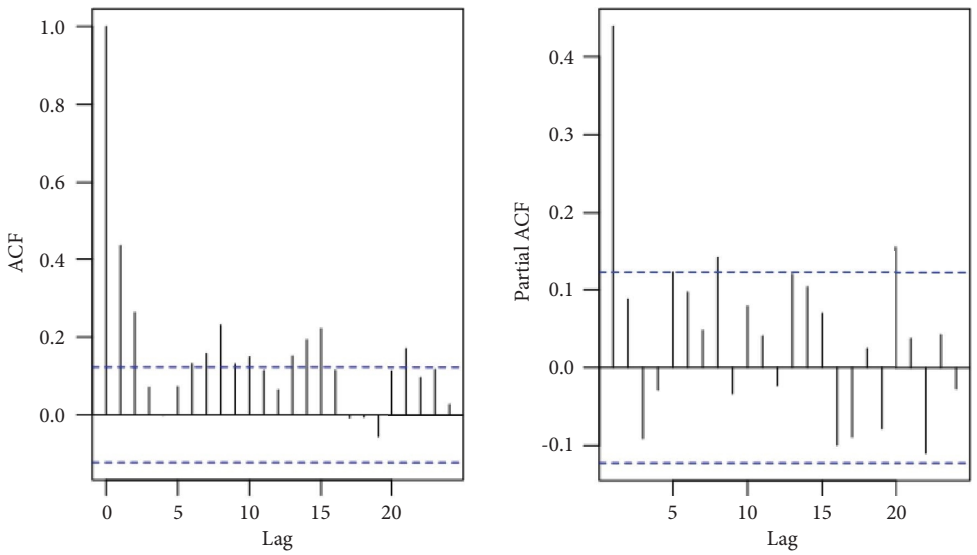

Figure 9: ACF and PACF plots of residuals of the GP model.

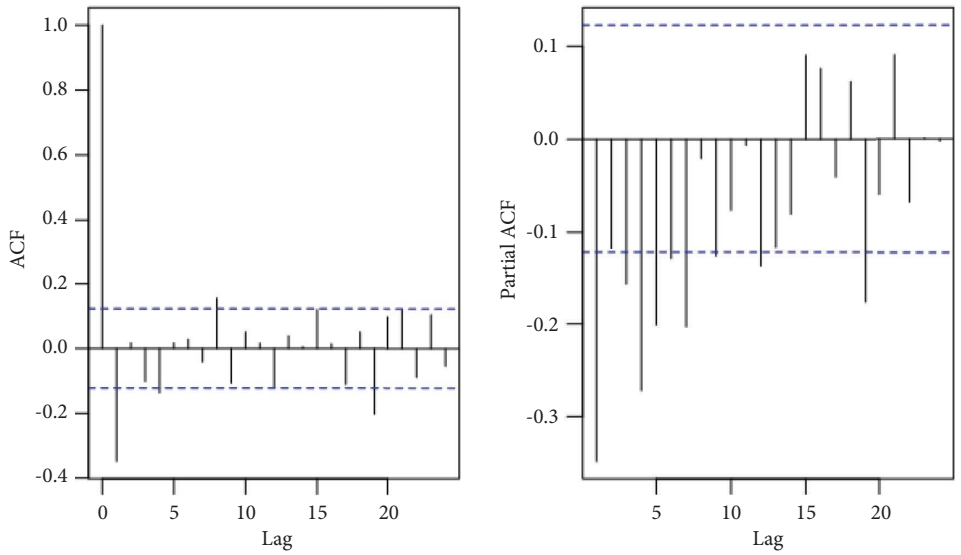

Figure 10: ACF and PACF plots of the differenced residuals of the GP regression model.

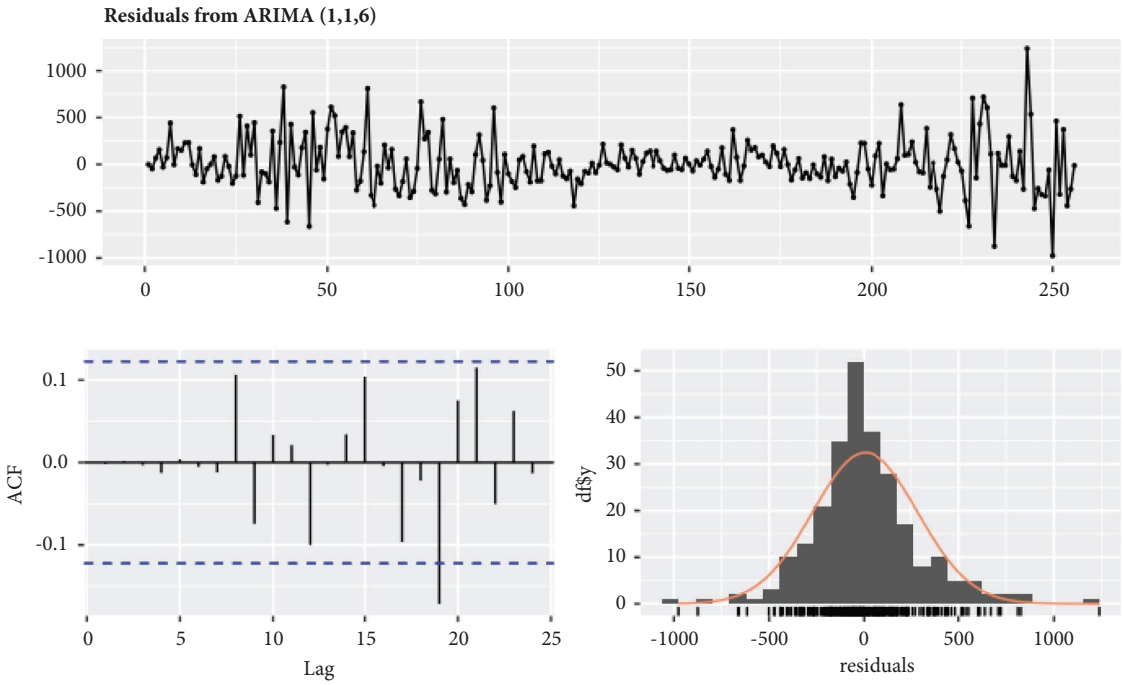

Figure 11: Residual analysis of ARIMA (1, 1, 6).

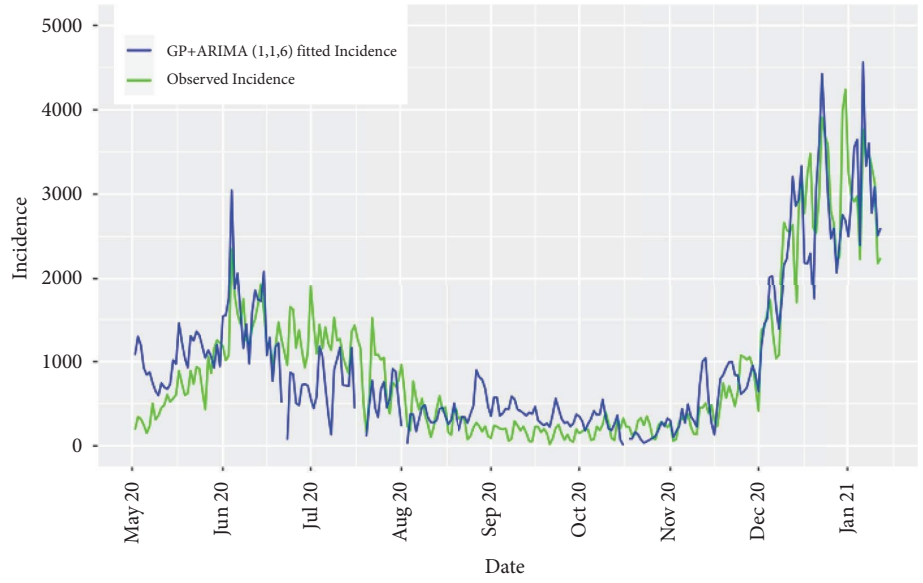

Figure 12: Actual and fitted values of GP + ARIMA (1, 1, 6) COVID-19 incidence.

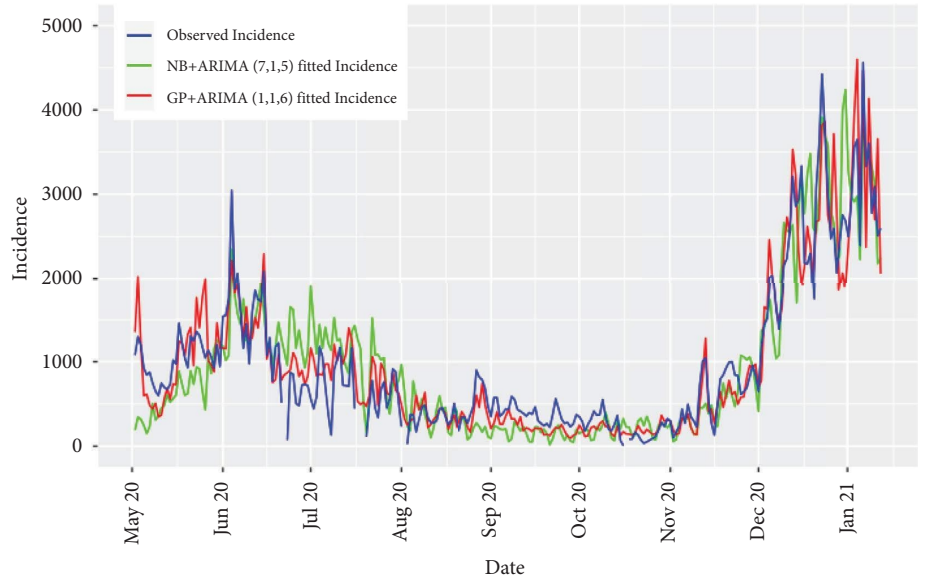

Figure 13: Comparison between observed incidence, fitted values of incidence by NB + ARIMA (7, 1, 5), and fitted values of incidence by GP + ARIMA (1, 1, 6).

Table 3: Estimates of generalised Poisson regression model for daily counts of COVID-19 incidence.

	Estimate	Std. error	z value	Pr(>|z|)
ϕ	1.6049	0.0216	74.170	0.0000***
Intercept	50.456	7.5623	6.672	0.0000***
Max humidity at lag 1	−0.0179	0.0049	−3.613	0.0000***
SO_2 at lag 1	−0.0500	0.0098	−5.075	0.0000***
$PM_{2.5}$ at lag 1	0.0617	0.0183	3.373	0.0000***
PM_{10} at lag 1	−0.0210	0.0065	−3.231	0.0012*
Min temperature at lag 2	0.0723	0.0217	3.331	0.0008***
Min humidity at lag 2	−0.0076	0.0028	−2.731	0.0063*
AmbRelHum at lag 2	0.0169	0.0037	4.524	0.0000***
Max temperature at lag 3	−0.0533	0.0215	−2.470	0.0135·
Min temperature at lag 3	0.0823	0.0240	3.422	0.0006***
Min humidity at lag 3	−0.0084	0.0038	−2.217	0.0266·
$PM_{2.5}$ at lag 3	0.0782	0.0177	4.405	0.0000***
PM_{10} at lag 3	−0.0266	0.0063	−4.177	0.0000***
Max temperature at lag 4	−0.0561	0.0200	−2.794	0.0052**
Min temperature at lag 4	0.0711	0.0229	3.094	0.0019**
NO at lag 4	−0.0211	0.0098	−2.146	0.0318·
$PM_{2.5}$ at lag 4	0.0538	0.0182	2.954	0.0031*
PM_{10} at lag 4	−0.0198	0.0065	−3.048	0.0023*
PM_Load 10 at lag 4	−0.0653	0.0196	−3.318	0.0009***
SolarRadiation at lag 4	−0.0014	0.0004	−3.082	0.0020*
Ambpressure at lag 4	−0.0426	0.0072	−5.860	0.0000***
Max humidity at lag 5	−0.0187	0.0051	−3.667	0.0002***
NO_2 at lag 5	0.0565	0.0131	4.283	0.0000***
NO_X at lag 5	−0.0450	0.0100	−4.486	0.0000***
AmbRelHum at lag 5	0.0189	0.0039	4.838	0.0000***
Min temperature at lag 6	0.0693	0.0177	3.904	0.0000***
Min humidity at lag 6	−0.0097	0.0028	−3.432	0.0006***
SO_2 at lag 6	−0.0528	0.0098	−5.393	0.0000***
$PM_{2.5}$ at lag 6	0.06387	0.0178	3.573	0.0003***
PM_{10} at lag 6	−0.0241	0.0063	−3.770	0.0002***
AmbWspeed at lag 6	0.1136	0.0264	4.291	0.0000***
SolarRadiation at lag 6	−0.0013	0.0004	−2.916	0.0035*

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1. Names of linear predictors: loglink (meanpar), logloglink (dispind). Log-likelihood: −1773.782 on 458 degrees of freedom. Number of fisher scoring iterations: 16.

4. Conclusion

The impact of pollution factors and meteorological variables on daily COVID-19 incidences in the Western Cape Province of South Africa is explored in this study. By taking into consideration the lags of the incubation period for COVID-19 disease, the study was able to model COVID-19 incidence counts using negative binomial and generalised Poisson regression models. The fitted negative binomial and generalised Poisson regression model residuals were autocorrelated. To capture the unexplained patterns that existed in the residuals, the ARIMA method was used. The residuals of the negative binomial regression model were fitted with an ARIMA (7, 1, 5), while the residuals of the generalised Poisson regression model were fitted with an ARIMA (1, 1, 6). The two models NB + ARIMA (7, 1, 5) and GP + ARIMA (1, 1, 6) were able to capture some of the trends found in the original series of incidence throughout the study period.

Based on the minimum RMSE, GP + ARIMA (1, 1, 6) was selected as the optimal model for COVID-19 incidence and was used to investigate the association between COVID-19 incidence with pollution and meteorological variables. The results revealed that $PM_{2.5}$, minimum temperature, NO_2, ambient relative humidity, and ambient wind speed at various lags were positively associated with COVID-19 incidence while maximum relative humidity, SO_2, PM_{10}, minimum relative humidity, maximum temperature, NO, PM load, solar radiation, and NO_X at various lags suggested a negative association with COVID-19 incidence. Ambient wind direction and temperature showed a nonsignificant association with COVID-19 at all lags. The positively associated variables can potentially enhance the risk of COVID-19 transmission while the negatively associated variables can control the risk of COVID-19 transmission. This study has supported the hypothesis that air pollution and meteorological variables impact COVID-19 transmission. Moreover, the findings of this study might be useful for future studies in other provinces and countries with similar meteorological and air pollution conditions.

This study, however, has some shortcomings. For instance, it does not account for confounding factors including population movements, population density, potential seasonal impacts, virus mutation, and public health interventions. This limited the ability to accurately measure the impact of meteorological variables and air pollution parameters on COVID-19 transmission. Second, performing stepwise regression, especially with a large number of lagged variables, has its limitations. Stepwise regression involves iteratively adding or removing variables based on certain criteria, such as significance levels or model fit statistics, which can lead to variable selection bias and inflated type I error rates. This process may indeed penalize regression estimates, particularly if not conducted with caution. Stepwise regression tends to select variables that best fit the sample data, potentially leading to overfitting and poor out-of-sample prediction performance. Stepwise procedures can be sensitive to outliers and multicollinearity, potentially biasing coefficient estimates and inflating standard errors. In addition, stepwise regression does not account for the uncertainty introduced by variable selection, leading to overly optimistic assessments of model performance and coefficient significance. Consequently, while stepwise regression can aid in model simplification, it is essential to interpret the results cautiously.

This study was conducted by analyzing data from the initial phase of the COVID-19 pandemic, which was influenced by changes in people's behavior and government-imposed containment measures. As a result, all important factors that may influence COVID-19 transmission should be identified through a comprehensive study and incorporated into the model to reduce any inconsistencies between the actual and fitted series. Another study should be undertaken utilizing the most current data and the results compared with the results of this study.

Data Availability

The data used to support the findings of this study are available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their heartfelt gratitude to the faculty and staff at AIMS Rwanda for their invaluable support and guidance throughout the funding for this research. This research, titled "Modeling the Impact of Air Pollution and Meteorological Variables on COVID-19 Transmission in Western Cape, South Africa," was presented as the culmination of the academic journeys at the African Institute for Mathematical Sciences, AIMS Rwanda, funded by the Next Einstein Initiative Scholarship [57]. The authors extend their sincere appreciation to thesis advisors for their mentorship and insightful feedback, significantly contributing to the development and refinement of this research. The academic environment at AIMS Rwanda provided us with an enriching experience. The authors are thankful for the opportunities to share and discuss their findings with the academic community. As the authors embark on the journey to publish this paper, they acknowledge the pivotal role AIMS Rwanda played in shaping research skills and fostering a passion for scientific inquiry. This study was funded by the African Institute for Mathematical Sciences.

References

[1] I. Fahmi, *World Health Organization Coronavirus Disease 2019 (Covid-19) Situation Report*, DroneEmprit, Jakarta, Indonesia, 2019.

[2] M. A. Lauxmann, N. E. Santucci, and A. M. Autrán-Gómez, "The sars-cov-2 coronavirus and the covid-19 outbreak," *International Brazilian Journal of Urology*, vol. 46, no. suppl 1, pp. 6–18, 2020.

[3] D. Cucinotta and M. Vanelli, "Who declares covid-19 a pandemic," *Acta BioMedica: Atenei Parmensis*, vol. 91, no. 1, pp. 157–160, 2020.

[4] WHO, *Modes of Transmission of Virus Causing Covid-19: Implications for Ipc Precaution Recommendations: Scientific Brief, 29 March 2020. Technical Report*, World Health Organization, Geneva, Switzerland, 2020.

[5] S. Lolli, Y.-C. Chen, S.-H. Wang, and G. Vivone, "Impact of meteorological conditions and air pollution on covid-19 pandemic transmission in Italy," *Scientific Reports*, vol. 10, no. 1, pp. 16213–16215, 2020.

[6] CDCP, *Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (Covid-19)*, National Center for Immunization and Respiratory Diseases (U.S.). Division of Viral Diseases, Atlanta, GA, USA, 2020.

[7] JHVDB, "John hopkins virus dashboard," 2022, https://coronavirus.jhu.edu/map.html.

[8] WHO, *Weekly Bulletin on Outbreak and Other Emergencies: Week 20: 11-17 May 2020. Weekly Bulletin On Outbreak And Other Emergencies*, WHO, Geneva, Switzerland, 2020.

[9] A. Haileamlak, "Covid-19 pandemic status in africa," *Ethiopian journal of health sciences*, vol. 30, no. 5, pp. 643-644, 2020.

[10] D. R. Silva, V. P. Viana, A. M. Müller, F. P. Livi, and P. de T. R. Dalcin, "Respiratory viral infections and effects of meteorological parameters and air pollution in adults with respiratory symptoms admitted to the emergency room," *Influenza and other respiratory viruses*, vol. 8, no. 1, pp. 42–52, 2014.

[11] Y. Iha, T. Kinjo, G. Parrott, F. Higa, H. Mori, and J. Fujita, "Comparative epidemiology of influenza a and b viral infection in a subtropical region: a 7-year surveillance in okinawa, Japan," *BMC Infectious Diseases*, vol. 16, no. 1, pp. 650–658, 2016.

[12] H. Li, X.-L. Xu, Da-W. Dai, Z.-Yu Huang, Z. Ma, and Y.-J. Guan, "Air pollution and temperature are associated with increased covid-19 incidence: a time series study," *International Journal of Infectious Diseases*, vol. 97, pp. 278–282, 2020.

[13] M. Lipsitch, *Seasonality of Sars-Cov-2: Will Covid-19 Go Away on its Own in Warmer Weather*, Center for communicable disease dynamics, Boston, MA, USA, 2020.

[14] M. Jerrett, "The death toll from air-pollution sources," *Nature*, vol. 525, no. 7569, pp. 330-331, 2015.

[15] Y. Cui, Z.-F. Zhang, J. Froines et al., "Air pollution and case fatality of sars in the people's Republic of China: an ecologic study," *Environmental Health*, vol. 2, no. 1, pp. 15–5, 2003.

[16] X. Wu, R. C. Nethery, M. Benjamin Sabath, D. Braun, and F. Dominici, "exposure to air pollution and covid-19 mortality in the united states: a nationwide cross-sectional study," *Medrxiv*, 2020.

[17] Y. Liu, Z. Ning, Yu Chen et al., "Aerodynamic analysis of sars-cov-2 in two wuhan hospitals," *Nature*, vol. 582, no. 7813, pp. 557–560, 2020.

[18] N. Ali and F. Islam, "The effects of air pollution on COVID-19 infection and mortality—a review on recent evidence," *Frontiers in Public Health*, vol. 8, 2020.

[19] J. S. M. Peiris and Yi Guan, "Severe acute respiratory syndrome," *Nature medicine*, vol. 10, no. S12, pp. S88–S97, 2004.

[20] M. S. Islam, M. Rahman, T. R. Tusher, S. Roy, and M. A. Razi, "Assessing the relationship between covid-19, air quality, and meteorological variables: a case study of dhaka city in Bangladesh," *Aerosol and Air Quality Research*, vol. 21, no. 6, Article ID 200609, 2021.

[21] Z. Zhang, T. Xue, and X. Jin, "Effects of meteorological conditions and air pollution on covid-19 transmission: evidence from 219 Chinese cities," *The Science of the Total Environment*, vol. 741, 2020.

[22] O. S. Makinde, B. M. Oseni, A. O. Adepetun, O. O. Olusola-Makinde, and G. Jacob Abiodun, "The significance of daily incidence and mortality cases due to covid-19 in some african countries," in *Data Science for COVID-19*, pp. 667–680, Elsevier, Amsterdam, Netherlands, 2022.

[23] A. K. Langat, J. K. Mutinda, S. M. Mwalili, and L. N. Kazembe, "Covid-19 impact analysis: assessing african sectors-commodity, service, manufacturing, and education using mixed model approach," *Asian Journal of Probability and Statistics*, vol. 25, no. 4, pp. 43–55, 2023.

[24] H. Cao, B. Li, T. Gu, X. Liu, K. Meng, and L. Zhang, "Associations of ambient air pollutants and meteorological factors with covid-19 transmission in 31 Chinese provinces: a time series study," *Inquiry: The Journal of Health Care Organization, Provision, and Financing*, vol. 58, 2021.

[25] I. Diouf, S. Sy, H. Senghor et al., "Potential contribution of climate conditions on covid-19 pandemic transmission over west and north african countries," *Atmosphere*, vol. 13, no. 1, p. 34, 2021.

[26] S. Ogunjo, O. Olaniyan, C. F. Olusegun, F. Kayode, D. Okoh, and G. Jenkins, "The role of meteorological variables and aerosols in the transmission of covid-19 during harmattan season," *GeoHealth*, vol. 6, no. 2, 2022.

[27] J. Yuan, Yu Wu, W. Jing et al., "Non-linear correlation between daily new cases of covid-19 and meteorological factors in 127 countries," *Environmental Research*, vol. 193, 2021.

[28] N. Ali and F. Islam, "The effects of air pollution on covid-19 infection and mortality—a review on recent evidence," *Frontiers in Public Health*, vol. 8, 2020.

[29] M. M. Sahoo, "Significance between air pollutants, meteorological factors, and covid-19 infections: probable evidences in India," *Environmental Science and Pollution Research*, vol. 28, no. 30, pp. 40474–40495, 2021.

[30] Y. Jiang, X.-J. Wu, and Y.-J. Guan, "Effect of ambient air pollutants and meteorological variables on covid-19 incidence," *Infection Control and Hospital Epidemiology*, vol. 41, no. 9, pp. 1011–1015, 2020.

[31] Y. Liang, L. Fang, H. Pan et al., "Pm 2.5 in beijing–temporal pattern and its association with influenza," *Environmental Health*, vol. 13, no. 1, pp. 102–108, 2014.

[32] S. Aggarwal, S. Balaji, T. Singh et al., "Association between ambient air pollutants and meteorological factors with sars-cov-2 transmission and mortality in India: an exploratory study," *Environmental Health*, vol. 20, no. 1, pp. 120–213, 2021.

[33] B. Bochenek, M. Jankowski, M. Gruszczynska et al., "Impact of meteorological conditions on the dynamics of the covid-19 pandemic in Poland," *International Journal of Environmental Research and Public Health*, vol. 18, no. 8, p. 3951, 2021.

[34] H. Tegally, E. Wilkinson, M. Giovanetti et al., "Detection of a sars-cov-2 variant of concern in South Africa," *Nature*, vol. 592, no. 7854, pp. 438–443, 2021.

[35] Nick roux, "Wikipedia file," 2022, https://commons.wikimedia.org/wiki/File.

[36] Wikipedia, "Wikipedia file," 2020, https://en.wikipedia.org/wiki/Western_Cape.

[37] Wcvdb, "Western cape covid-19 dashboard," 2022, https://coronavirus.westerncape.gov.za/covid-19-dashboard.

[38] C. Marivate, "Covid-19 data repository," 2020, https://github.com/dsfsi/covid19za.

[39] saaqis, "South african weather service air quality," 2020, https://saaqis.environment.gov.za/.

[40] L. Paulo Fávero, R. de Freitas Souza, P. Belfiore, H. L. Corrêa, and M. F. C. Haddad, "Count data regression analysis: concepts, overdispersion detection, zero-inflation identification, and

applications with r," *Practical Assessment, Research and Evaluation*, vol. 26, no. 1, p. 13, 2021.

[41] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[42] G. J. Abiodun, O. S. Makinde, A. M. Adeola et al., "A dynamical and zero-inflated negative binomial regression modelling of malaria incidence in limpopo province, South Africa," *International Journal of Environmental Research and Public Health*, vol. 16, no. 11, p. 2000, 2019.

[43] O. S. Makinde, A. M. Adeola, G. J. Abiodun, O. O. Olusola-Makinde, and A. Alejandro, "Comparison of predictive models and impact assessment of lockdown for covid-19 over the United States," *Journal of Epidemiology and Global Health*, vol. 11, no. 2, p. 200, 2021.

[44] deL. Jan, "Introduction to akaike (1973) information theory and an extension of the maximum likelihood principle," in *Breakthroughs in Statistics*, pp. 599–609, Springer, Berlin, Germany, 1992.

[45] F. Famoye, "Restricted generalized Poisson regression model," *Communications in Statistics- Theory and Methods*, vol. 22, no. 5, pp. 1335–1354, 1993.

[46] PoC. Consul and F. Famoye, "Generalized Poisson regression model," *Communications in Statistics- Theory and Methods*, vol. 21, no. 1, pp. 89–109, 1992.

[47] A. Pankratz, *Forecasting with Dynamic Regression Models*, John Wiley & Sons, Hoboken, NJ, USA, 2012.

[48] D. R. McQ. Allan and C.-L. Tsai, *Regression and Time Series Model Selection*, World Scientific, Singapore, 1998.

[49] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)," *Geoscientific Model Development Discussions*, vol. 7, no. 1, pp. 1525–1534, 2014.

[50] S. Sangkham, S. Thongtip, and P. Vongruang, "Influence of air pollution and meteorological factors on the spread of covid-19 in the bangkok metropolitan region and air quality during the outbreak," *Environmental Research*, vol. 197, 2021.

[51] N. Wardhani, H. Gani, S. Zuhriyah, H. Gani, and E. Vidyarini, "A correlation method for meteorological factors and air pollution in association to covid-19 pandemic in the most affected city in Indonesia," *ILKOM Jurnal Ilmiah*, vol. 13, no. 3, pp. 195–205, 2021.

[52] P. Sharma, A. K. Singh, B. Agrawal, and A. Sharma, "Correlation between weather and covid-19 pandemic in India: an empirical investigation," *Journal of Public Affairs*, vol. 20, no. 4, 2020.

[53] S. Baweja, S. Thangariyal, A. Rastogi, A. Tomar, and A. Bhadoria, "Impact of temperature and sunshine duration on daily new cases and death due to covid-19," *Journal of Family Medicine and Primary Care*, vol. 9, no. 12, p. 6091, 2020.

[54] J. Yuan, Yu Wu, W. Jing et al., "Association between meteorological factors and daily new cases of covid-19 in 188 countries: a time series analysis," *Science of the Total Environment*, vol. 780, 2021.

[55] G. Isaia, H. Diémoz, F. Maluta et al., "Does solar ultraviolet radiation play a role in covid-19 infection and deaths? an environmental ecological study in Italy," *Science of the Total Environment*, vol. 757, 2021.

[56] K. Sharun, R. Tiwari, and K. Dhama, "Covid-19 and sunlight: impact on sars-cov-2 transmissibility, morbidity, and mortality," *Annals of medicine and surgery (2012)*, vol. 66, 2021.

[57] John Kamwele Mutinda, *African institute for Mathematical Sciences*, (Aims), rwanda, Kigali, Rwanda.