Journal of Biosciences and Medicines, 2022, 10, 77-86
https://www.scirp.org/journal/jbm
ISSN Online: 2327-509X
ISSN Print: 2327-5081

Modeling and Forecasting of SARS CoV-2 Cases in Sierra Leone

Sallieu Kabay Samura[1*], Theresa Ruba Koroma[2], Abdul A. Kamara[1]

[1]Department of Mathematics and Statistics, Fourah Bay College, University of Sierra Leone, Freetown, Sierra Leone
[2]Department of Internal Medicine, Faculty of Clinical Sciences, College of Medicine and Allied Health Sciences, University of Sierra Leone, Freetown, Sierra Leone
Email: *ssallieu@yahoo.com

How to cite this paper: Samura, S.K., Koroma, T.R. and Kamara, A.A. (2022) Modeling and Forecasting of SARS CoV-2 Cases in Sierra Leone. *Journal of Biosciences and Medicines*, **10**, 77-86.
https://doi.org/10.4236/jbm.2022.104008

Received: February 9, 2022
Accepted: April 5, 2022
Published: April 8, 2022

Abstract

Severe acute respiratory syndrome coronavirus-2 (SARS CoV-2) has been a global threat spreading in Sierra Leone, and many studies are being conducted using various Statistical models to predict the probable evolution of this pandemic. In this paper, we use the autoregressive integrated moving average (ARIMA) model with the aim of forecasting the cumulative confirmed cases of SARS CoV-2 in Sierra Leone. The Akaike Information Criterion (AIC) was applied to the training data as a criterion method to select the best model. In addition, the statistical measure RMSE and MAPE were utilized for testing this data, and the model with the minimum RMSE and MAPE was selected for future forecasting. ARIMA (3, 2, 1) was confirmed to be the optimal model based on the lowest AIC value. This model was then applied to study the trend of SARS CoV-2 from 1st February 2022 to 30th February 2022. The result shows that incidence of SARS CoV-2 from 1st February 2022 to 30th February 2022, increasing growth steep in Sierra Leone (7718.629, 95% confidence limit of 6785.985 - 8651.274).

Keywords

ARIMA Model, SARS Cov-2, Stationarity, Forecast

1. Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS CoV-2) is a novel coronavirus that broke out in Wuhan China in December 2019 and has rapidly spread across the world causing a global pandemic [1] [2]. To date, SARS CoV-2 is the seventh coronavirus known to affect human beings. It affects the lower respiratory tract and causes symptoms ranging from mild fever, cough and sore throat to severe and fatal complications including acute respiratory distress syndrome (ARDS), severe pneumonia, septic shock, pulmonary edema, hypercoagulable state, other organ failure and subsequent death [3] [4]. Patients with underlying comorbidities like diabetes mellitus, cardiovascular diseases, malignancies, chronic respiratory diseases and elderly people are more prone to developing complications of SARS CoV-2.

According to the Worldometer as of January 29th, 2022, there had been 372,153,572 confirmed cases, 5,672,345 reported deaths and 293,778,361 recovered SARS CoV-2 cases worldwide [5]. The spread of this disease has been a growing public health concern as it affects and poses significant challenges to a country's economic, political and social development. Several nations have tried to curtail this spread by imposing strict hand hygiene and imposing national lockdown. Unfortunately, given their large genetic diversity, frequent genome recombination, multiple viral strains with identified genetic polymorphisms, complex disease manifestation, and multiple routes of transmission of SARS CoV-2, control measures have not been very successful [6].

Sierra Leone reported its first case in March of 2020 [7] and since then there have been 7608 confirmed cases of SARS CoV-2 with 125 deaths [8]. As of 26 January 2022, a total of 1,409,313 vaccine doses have been administered [8]. A recent cross-sectional, nationally representative, age-stratified serosurvey on SARS-CoV-2 antibody prevalence in Sierra Leone shows that overall weighted seroprevalence was 2.6% (95% CI 1.9 - 3.4) which is 43 times higher than the reported number of cases [9]. Despite this relatively low rate of infection and spread of SARS-CoV-2 in Sierra Leone, there is still a lot of uncertainty regarding this ever-changing coronavirus. The healthcare system in this country is very frail and unequipped for large number of admissions at a go and with an economy that is largely dependent on imports and exports; long-term border closure is not feasible.

Therefore to help predict the trajectory of the disease and for short-term forecasting of new cumulative confirmed cases, we utilized a univariate autoregressive integrated moving average (ARIMA). The ARIMA models have been successfully applied to predict the incidence of infectious diseases, such as influenza mortality [10], malaria incidence [11], as well as other infectious diseases [12] [13]. This model is vital for understanding and estimating the disease progression in Sierra Leone, which will help inform policy planning regarding curtailing further spread and future containment measures.

2. Methods

2.1. Data Source

The data for this study consists of confirmed SARS CoV-2, cases per day from 13th March 2020 to 30th January 2022. The daily SARS CoV-2, cases were obtained from Our World in Data, an official website for all SARS CoV-2, (https://covid19.who.int/region/afro/country/sl). We use R statistical software to analyze the SARS CoV-2, data.

2.2. Unit Root Test

Before estimating the parameters for the ARIMA model, the data were tested for stationarity using the Augmented Dickey-Fuller (ADF) test, for which the null hypothesis H_0 of the time series is said to be non-stationary. The result of the ADF test suggested that the time-series data was non-stationary (p > 0.05). After applying the second difference, *i.e.*, $d(0)$, the p-value obtained was less than the significance level ($p < 0.05$) and the statistical ADF is lower than any of the critical values, so the null hypothesis was rejected.

2.3. The Model

The autoregressive integrated moving average (ARIMA) model, is a generalization of the ARMA model with non-stationary series. ARIMA is non-stationary means that it has non-constant mean and variance over time. The integrated part refers to a differencing initial step, which can be applied to eliminate the non-stationarity of the series. An ARIMA model is unequivocal by its three components:

- *Auto regression* (*AR*) model is the model which represents a variable that regresses on its lagged, or prior, values.
- *Integrated* (*I*) shows the differencing of basic observations so that the time series may be stationary.
- *Moving average* (*MA*) provides the docility between an observation and a residual from the MA model for lag observations.

The autoregressive time series regression model of order *p*, signified by AR(*p*) is given by

$$\begin{cases} x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t \\ \varphi_p \neq 0 \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, \quad E(\varepsilon_t \varepsilon_s) = 0, \quad s \neq t \\ E(x_s \varepsilon_t) = 0, \quad \forall s < t \end{cases}$$

where $\{\varphi_t\}, i = 1, \cdots, p$ is the model parameter, $\{\varepsilon_t\}, t = 1, \cdots, p$ is a normally distributed random process with mean 0 and a constant variance σ_ε^2 which is assumed to be independent of all process values.

White noise series properties with mean 0 and variance σ_ε^2 are moving averages, with order *q* expressed as MA (*q*). The weighted linear sum of previous forecast errors is given by

$$\begin{cases} x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_p \varepsilon_{t-q} \\ \theta_p \neq 0 \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, \quad E(\varepsilon_t \varepsilon_s) = 0, \quad s \neq t \end{cases}$$

where $\{\theta_t\}, i = 1, \cdots, p$ is the model parameter, $\{\varepsilon_t\}, t = 1, \cdots, p$ is a normally distributed random process with mean 0 and a constant variance σ_ε^2 which is assumed to be independent of all process values.

The ARMA (*p*, *q*) model composes of two main polynomials which are AR(*p*) and MA (*q*). It is expressed thus:

$$\begin{cases} x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \varphi_p \neq 0,\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, \quad E(\varepsilon_t \varepsilon_s) = 0, \quad s \neq t \\ E(x_s \varepsilon_t) = 0, \quad \forall s < t \end{cases}$$

where $x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$ is $\Phi(B)\Delta^d x_t = \Theta(B)\varepsilon_t$. $\{\varphi_t\}, i = 1, \cdots, p$ and $\{\theta_t\}, i = 1, \cdots, p$ are the model parameters, $\{\varepsilon_t\}, t = 1, \cdots, p$ is a normally distributed random process with mean 0 and a constant variance σ_ε^2 which is assumed to be independent of all process values.

The ARIMA (*p*, *d*, *q*) model is a widely used statistical method used in stationary time-series analysis such as forecasting. To build such a model, the primary step is to investigate whether the statistical stationery of a time series can be satisfied or not. Then, the next phase is estimating the numerical values of *p* and *q* parameters for AR and MA models. Thus, the essential idea of the ARIMA model is based on the assumption that the predicted value of the variable x_t is generated from a linear equation of several previous observations with random errors. A process x_t is an ARIMA (*p*, *d*, *q*) when it satisfies the form

$$\begin{cases} \Phi(B)\Delta^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, \quad E(\varepsilon_t \varepsilon_s) = 0, \quad s \neq t \\ E(x_s \varepsilon_t) = 0, \quad \forall s < t \end{cases}$$

where $\phi(B)$ and $\theta(B)$ are polynomial operators. $\Delta^d x_t = (1-B)^d x_t$, for $d \geq 1$, where $\Delta = 1 - B$ is the difference operator.

2.4. Performance Measures

To evaluate the prediction models, we use the following statistical measures.

Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k - \hat{x}_k)^2}$$

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{100}{N}\sum_{k=1}^{N}\left|\frac{x_k - \hat{x}_k}{x_k}\right|$$

where x_k denotes actual value and $\hat{x}_k$ denotes the predicted value for the *k*th instance.

3. Results

Figure 1 shows a strong upward trend of SARS CoV-2, cases in Sierra Leone showing that the series is not stationary. This is confirmed by results of the unit

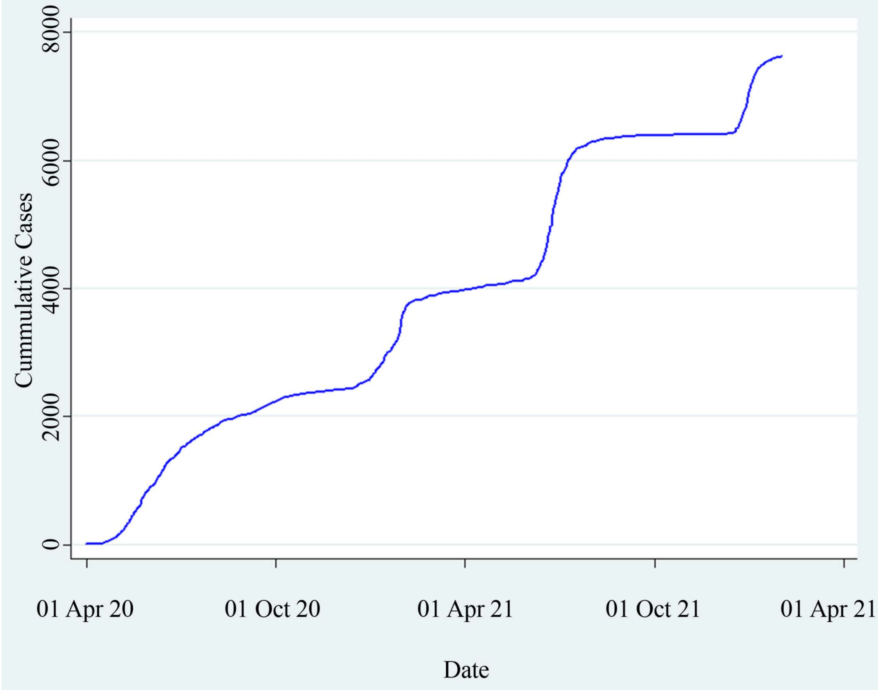

Figure 1. Cumulative confirmed cases of SARS CoV-2 from 3rd March 2020 to 31st January 2022.

root tests ADF as presented in **Table 1**, where the p-values are all greater than 5% level of significance. Thus, there is not enough evidence to reject the null hypothesis that the SARS CoV-2, series of Sierra Leone is nonstationary. Nonetheless, a second difference in the series made it stationary, as confirmed by the ADF.

The autocorrelation function (ACF) plot is also useful for identifying nonstationary time series. For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly. Therefore, differencing can help stabilize the mean of a time series by removing changes in the level of a time series, and therefore eliminating (or reducing) trend. Consequently, we will take a second difference in the data. The second differenced data are shown in **Figure 2**.

Residuals are useful for testing the model's suitability to capture the information in the data. The estimated autocorrelations between the residuals at various lags are depicted in **Figure 3**.

From the ACF and PACF graph (as shown in **Figure 2**) and the models trace summary table (**Table 2**), we were able to observe the following candidate models and also using the AICc model selection criterion, we detect that the ARIMA (3, 2, 1) with drift as the model with lowest AICc value.

The ARIMA (3, 2, 1) model predicts the number of cumulative confirmed cases over the next 30 days using the previously observed data as shown in **Table 3** with lower and upper confidence limits. Although the increasing trend is visible, the model has better performance.

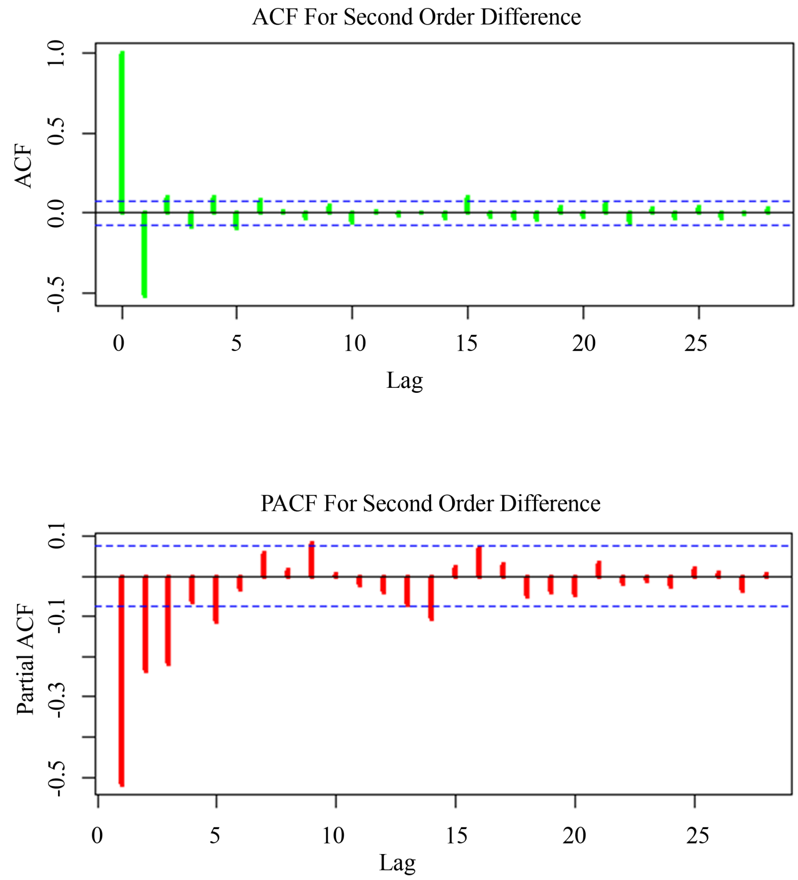

Figure 2. ACF and PACF plot for second-order differenced cumulative confirmed cases of SARS CoV-2.

Table 1. ADF unit root tests on log levels of variables.

Augmented Dickey-Fuller test for unit root, Number of obs. = 669						
Interpolated Dickey-Fuller						
	Test Statistics	1% Critical Value	5% Critical value	10% Critical Value		
Z(t)	−2.307	−3.960	−3.410	−3.120		
Mackinnon approximate p-value for Z(t) = 1.00						
D. Cum. Cases	Coef.	Std. Err.	t	p > |t|	[95% Conf. Interval]	
D.Cum. Cases						
L1.	−0.0030118	0.0013057	−2.31	0.021	−0.0055757	−0.004479
LD.	0.3847551	0.0350505	10.98	0.000	0.315932	0.4535783
L2D.	0.4267533	0.0350943	12.16	0.000	0.3578442	0.4956624
_trend	0.329497	0.0146564	2.25	0.025	0.0041712	0.617281
_cons	2.527277	1.043666	2.42	0.016	0.4779936	4.57656

Source: STATA software.

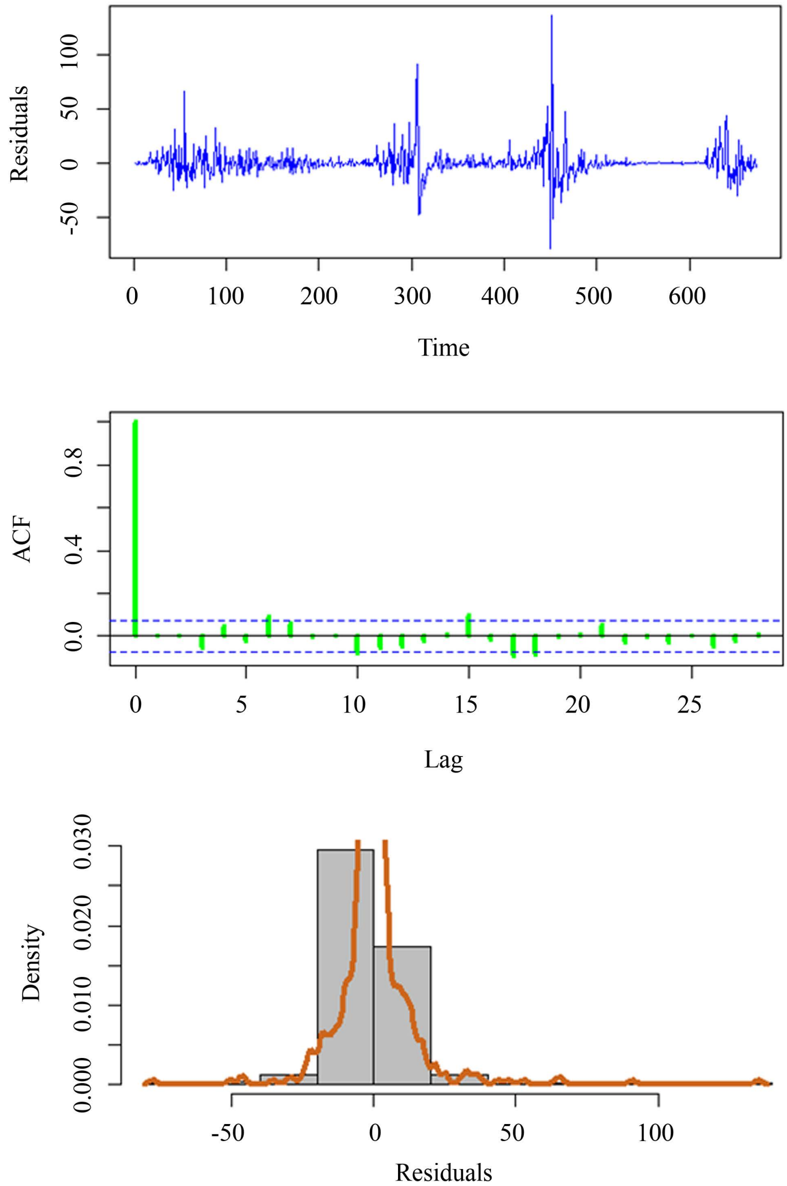

Figure 3. Residual plots form the ARIMA (3, 2, 1) model total confirmed cases of SARS CoV-2.

Table 2. AIC, MAPE and RMSE values for various ARIMA models applied for cumulative confirmed cases of SARS CoV-2.

Models	AIC	MAPE	RMSE
ARIMA (2, 2, 1)	5284.94	0.8542741	12.39312
ARIMA (3, 2, 1)	5283.07	0.8444274	12.35716
ARIMA (3, 2, 2)	5284.04	0.8432304	12.34766
ARIMA (2, 2, 2)	5284.42	0.8514826	12.36977
ARIMA (1, 2, 1)	5283.05	0.8536747	12.39417
ARIMA (1, 2, 0)	5355.04	0.9176175	13.10052

Table 3. Performance of ARIMA (3, 2, 1) model with 80% and 95% CI.

Date	Forecast	80% Lower	80% Upper	95% Lower	95% Upper
2/01/2022	7626.395	7610.535	7642.254	7602.139	7650.650
2/02/2022	7629.309	7603.505	7655.113	7589.845	7668.773
2/03/2022	7632.785	7595.976	7669.595	7576.490	7689.081
2/04/2022	7635.893	7587.686	7684.100	7562.167	7709.619
2/05/2022	7639.095	7578.075	7700.115	7545.773	7732.418
2/06/2022	7642.247	7567.443	7717.050	7527.844	7756.649
2/07/2022	7645.443	7555.903	7734.983	7508.503	7782.383
2/08/2022	7648.622	7543.530	7753.715	7487.897	7809.347
2/09/2022	7651.807	7530.345	7773.269	7466.046	7837.568
2/10/2022	7654.987	7516.385	7793.589	7443.013	7866.961
2/11/2022	7658.170	7501.685	7814.654	7418.847	7897.492
2/12/2022	7661.352	7486.277	7836.426	7393.598	7929.105
2/13/2022	7664.534	7470.186	7858.882	7367.304	7961.763
2/14/2022	7667.716	7453.436	7881.996	7340.002	7995.429
2/15/2022	7670.898	7436.047	7905.749	7311.724	8030.072
2/16/2022	7674.080	7418.039	7930.121	7282.499	8065.661
2/17/2022	7677.262	7399.430	7955.094	7252.354	8102.170
2/18/2022	7680.444	7380.235	7980.653	7221.314	8139.574
2/19/2022	7683.626	7360.471	8006.782	7189.402	8177.850
2/20/2022	7686.808	7340.150	8033.467	7156.639	8216.977
2/21/2022	7689.990	7319.285	8060.696	7123.045	8256.936
2/22/2022	7693.173	7297.889	8088.456	7088.638	8297.707
2/23/2022	7696.355	7275.972	8116.737	7053.435	8339.274
2/24/2022	7699.537	7253.546	8145.528	7017.452	8381.621
2/25/2022	7702.719	7230.620	8174.818	6980.706	8424.732
2/26/2022	7705.901	7207.204	8204.598	6943.209	8468.593
2/27/2022	7709.083	7183.306	8234.860	6904.976	8513.190
2/28/2022	7712.265	7158.935	8265.595	6866.020	8558.510
2/29/2022	7715.447	7134.099	8296.795	6826.352	8604.542
2/30/2022	7718.629	7108.806	8328.453	6785.985	8651.274

4. Discussion and Conclusions

The ARIMA model is one of the most widely used time-series forecasting techniques because of its structured modeling basis and acceptable forecasting performance [14]. In this paper, we applied an ARIMA (*p*, *d*, *q*) model to analyze the surveillance data of SARS CoV-2, infection in Sierra Leone. We have obtained an ARIMA model that closely fits the spread of SARS CoV-2, in Sierra Leone. According to the results above, the conducted model is reliable with high validity. Once a satisfactory model has been obtained, it can be used to forecast expected numbers of cases for a given number of future time intervals [15]. The forecast results suggest that the cumulative confirmed cases of SARS CoV-2, in Sierra Leone will experience strong growth in the next 30 days (22nd January 2020 to 19th February 2022).

As mentioned above, for adequate ARIMA modeling, a time series should be stationary with respect to mean and variance [16]. If the mean increases or decreases over time, or if the variance does, the series may need to be transformed to make it stationary, before being modeled. Otherwise, the prediction effect of the model will be poor. In order to improve the model, updating the forecasts is very important. A model without seasonal terms will need to be updated frequently. Confidence intervals that widen rapidly as time increases from the starting point of the forecasts also indicate a model that needs frequent updating. Generally speaking, there are two ways to implement the update. The model can be reapplied to the original series with extra observations added at the end to give forecasts based on a later starting point. Alternatively, a new model can be fitted to the longer series. This is probably preferable, since fitting a model is quick, especially when the old model is used as a guide, and it makes better use of the additional observations.

Government of Sierra Leone through the National SARS CoV-2. Emergency Operations Center (NACOVAC) can apply the forecasted trend of much more spread to make more informed decisions on the additional measures in place to curb the spread of the virus. Application of the model can also assist in studying the effectiveness of the lockdown on the spread of SARS CoV-2 in Sierra Leone.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F. and Tan, W. (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England Journal of Medicine*, **382**, 727-733. https://doi.org/10.1056/NEJMoa2001017

[2] Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. and Garry, R.F. (2020) The Proximal Origin of SARS-CoV-2. *Nature Medicine*, **26**, 450-452. https://doi.org/10.1038/s41591-020-0820-9

[3] Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C. and Agha, R. (2020) World Health Organization Declares Global Emergency: A Review of the 2019 Novel Coronavirus (COVID-19). *International Journal of Surgery*,

76, 71-76. https://doi.org/10.1016/j.ijsu.2020.02.034

[4] Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X. and Zhang, L. (2020) Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study. *The Lancet* (*London*, *England*), **395**, 507-513. https://doi.org/10.1016/S0140-6736(20)30211-7

[5] (Live). WMCVU. https://www.worldometers.info/coronavirus

[6] Duong, N.Q., Phuong Thao, L., Nhu Quynh, D.T., Binh, L.T., Ai Loan, C.T. and Hong Diem, P.T. (2020) Predicting the Pandemic COVID-19 Using ARIMA Model. *VNU Journal of Science*: *Mathematics—Physics*, **36**, 46-57. https://doi.org/10.25073/2588-1124/vnumap.4492

[7] Sierra Leone Confirms First Case of COVID-19. https://www.afro.who.int/news/sierra-leone-confirms-first-case-covid-19

[8] Sierra Leone. https://covid19.who.int/region/afro/country/sl

[9] Barrie, M.B., Lakoh, S., Kelly, J.D., Kanu, J.S., Squire, J., Koroma, Z., Bah, S., Sankoh, O., Brima, A., Ansumana, R., Goldberg, S.A., Chitre, S., Osuagwu, C., Maeda, J., Barekye, B., Numbere, T.W., Abdulaziz, M., Mounts, A., Blanton, C., Singh, T., Samai, M., Vandi, M.A. and Richardson, E.T. (2021) SARS-CoV-2 Antibody Prevalence in Sierra Leone, March 2021: A Cross-Sectional, Nationally Representative, Age-Stratified Serosurvey. *BMJ Glob Health*, **6**, e007271. https://doi.org/10.1136/bmjgh-2021-007271

[10] Reichert, T.A., Simonsen, L., Sharma, A., Pardo, S.A., Fedson, D.S. and Miller, M.A. (2004) Influenza and the Winter Increase in Mortality in the United States, 1959-1999. *American Journal of Epidemiology*, **160**, 492-502. https://doi.org/10.1093/aje/kwh227

[11] Gaudart, J., Touré, O., Dessay, N., *et al.* (2009) Modelling Malaria Incidence with Environmental Dependency in a Locality of Sudanese Savannah Area, Mali. *Malaria Journal*, **8**, 61. https://doi.org/10.1186/1475-2875-8-61

[12] Luz, P.M., Mendes, B.V., Codeço, C.T., Struchiner, C.J. and Galvani, A.P. (2008) Time Series Analysis of Dengue Incidence in Rio de Janeiro, Brazil. *American Journal of Tropical Medicine and Hygiene*, **79**, 933-939. https://doi.org/10.4269/ajtmh.2008.79.933

[13] Yi, J., Du, C.T., Wang, R.H. and Liu, L. (2007) Applications of Multiple Seasonal Autoregressive Integrated Moving Average (ARIMA) Model on Predictive Incidence of Tuberculosis. *Chinese Journal of Preventive Medicine*, **41**, 118-121.

[14] Box, G.E.P. and Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. Revised Edition, Holden Day, San Francisco. https://doi.org/10.1080/02664768900000007

[15] Haines, L.M., Munoz, W.P. and Van Gelderen, C.J. (1989) ARIMA Modelling of Birth Data. *Journal of Applied Statistics*, **16**, 55-67.

[16] Cheung, Y.-W. and Lai, K.S. (1995) Lag Order and Critical Values of the Augmented Dickey-Fuller Test. *Journal of Business & Economic Statistics*, **13**, 277-280. https://doi.org/10.1080/07350015.1995.10524601