Time Series Analysis and Forecast of COVID-19 Data in Southern African Countries

Harouna Sangaré and Abdou Fané

Abstract — In this paper, we begin by analyzing different time series COVID-19 data by exploring confirmed, death and recovered daily reports from January 22, 2020 to September 2nd, 2021 in the southern African counties. We then use the Prophet model to forecast the total numbers of infected cases from September 3rd, 2021 to December 1st, 2021 by providing graphics and by giving tables containing the total numbers of infected cases from November 27, 2021 to December 1st, 2021 for each country in the zone.

Keywords — COVID-19, data analysis, forecast, prophet model, time series.

I. INTRODUCTION

The respiratory virus coronavirus disease 2019 (COVID-19) that appeared in the city of Wuhan, China, in December 2019 has so far exhibited its capacity to produce severe outbreaks in restricted settings and traverse borders succeeding human movement patterns. In Africa, the first case was detected in Egypt from a non-national on February 15, 2020. The West African Economic and Monetary Union zone recorded its first case of confirmed acute respiratory disease due to coronavirus on March 2, 2020. Since then, countries throughout the zone have adopted stringent actions to protect the populations from the infection proliferation according to [1]. In their paper, they presented the current state of the coronavirus disease (COVID-19) in the Union (over the period from March 2, 2020, to June 20, 2020), with a focus on the first confirmed case, death and recovered history, subregional safeguards, preparedness among national governments, and socio-economic impacts and their results revealed that most of the first confirmed cases were imported, and the first death cases are mostly aged people with antecedent health issues.

Although many results concerning the COVID-19 data study are available, there are still many open problems about the study of disease spread since data continue to be recorded.

In this paper, we are interested in the evolution study of COVID-19 in southern African counties.

More precisely, we intend to investigate recent COVID-19 data and to provide fresh information by exploring different time series for each country in the southern African countries and by forecasting the total numbers of infected cases using the Prophet model. We think the forecast is an important factor while taking decision for Governments and private sectors.

Reference [2] are dedicated a chapter of their book to the Prophet model. In their book, they present the high-level math behind Prophet, how to create a basic model, how to apply log transformations to time-series data, how to add country holidays, and how to handle multivariate time-series data using add regressor.

From the COVID19 analytics R package developed by [3], we obtain live worldwide data from the novel CoronaVirus Disease originally reported in 2019, COVID-19, for a period of 590 days (from January 22, 2020 to September 2nd, 2021) as published by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) data repository https://github.com/CSSEGISandData/COVID-19, as well as, some basic analysis tools and functions to investigate these datasets.

Indeed, the authors (see [3]) present the COVID19.analytics package that allows users to access and analyze worldwide data from resources publicly available. they also introduced it focusing on its capabilities and presenting a particular study case where they describe how to deploy the COVID19.analytics Dashboard Explorer.

The paper is organized as follows. The methodology is described in Section 2. Section 3 is concerned by results, commentaries and remarks. The paper is ended by the conclusion Section 4.
II. METHODS

The first part of our studies focuses on the exploratory analysis of different time series (confirmed and death). We investigate these datasets by visualizing, at first, the daily cases of each country in the southern African zone and then the total numbers of different cases in comparative graphics.

Secondly, we forecast the total numbers of confirmed cases using the Prophet model which is a decomposable time series model with three main model components: trend, seasonality, and holidays ([2] and [4]). They are combined in the following equation:

\[ y(t) = g(t) + s(t) + h(t) + \epsilon_t \]

where \( g(t) \) is trend (linear/logistic), \( s(t) \) the periodic change/seasonality where, \( P \) is the consistent period we anticipate the time series to have. \( h(t) \) is the effect of holiday. For each holiday \( i \), let \( D_i \) be the set of past and future dates for that holiday, as shown here:

\[ Z(t) = [1_{(t \in D_1)}, \ldots, 1_{(t \in D_L)}] \]

and taking the following:

\[ h(t) = Z(t)\kappa. \]

As with seasonality, we use a prior \( \kappa \sim N(0, \sigma^2) \). 
\( \epsilon_t \) is the changes that are not adopted by the model.

Prophet is an open-source framework from Facebook used for framing and forecasting time series. It focuses on an additive model where nonlinear trends fit with daily, weekly, and yearly seasonality and additional holiday effects. Prophet is powerful at handling missing data and shifts within the trends and generally handles outliers well. It also allows user to accumulate exogenous variables to the model.

Many machine learning algorithms (machine learning Support Vector Machine (SVM), Prophet Forecasting Model, and Linear Regression Model) have been experimentally tested for the time series analysis of forecasting the predicted active, death and cured rate on novel COVID-19 pandemic data set by [5]. In their paper, Prophet Forecasting Model has been shown to be the best predictive method for predicting active rate, death rate and cured rate compared to the others.

We are going to generate forecasts from September 3\textsuperscript{rd}, 2021 to December 1\textsuperscript{st}, 2021 using Prophet model. R and R Studio software are used to generate graphs and excepted values.

III. RESULTS AND COMMENT

In this section, we begin by presenting an exploratory analysis of time series by visualizing the daily cases report for each country in the southern African area in Section A followed by the total numbers of cases in comparative graphics in Section B and we then provide a forecasting analysis of the total numbers of infected cases using the Prophet model in Section C.

A. Daily Cases Analytics

Graphics in Fig. 1 represent the numbers of confirmed cases per day for each country. Regarding these diagnostic plots, we observe that all these countries have reported their first case in March 2020. Angola reached the peak of daily cases on 2021-05-26 with a maximum of 405 cases. Botswana reached the peak of daily cases on 2021-08-02 with a maximum of 8530 cases. Lesotho reached the peak of daily cases on 2021-01-06 with a maximum of 931 cases. Malawi reached the peak on 2021-01-22 with a maximum of 1316 cases. Mozambique reached the peak of daily cases on 2021-07-28 with a maximum of 2460 cases. Namibia reached the peak of daily cases on 2021-06-30 with a maximum of 3268 cases. South Africa reached the peak of daily cases on 2021-07-03 with a maximum of 26485 cases. Zambia reached the peak of daily cases on 2021-06-24 with a maximum of 3594 cases. Zimbabwe reached the peak of daily cases on 2021-07-14 with a maximum of 3110 cases.

Graphics in Fig. 2 represent the numbers of deaths per day. These diagnostic plots show that all these countries have reported their first case in April 2020. Angola reached the peak of daily cases on 2021-05-17 with a maximum of 18 deaths. Botswana reached the peak of daily cases on 2021-08-12 with a maximum of 141 deaths. Lesotho reached the peak of daily cases on 2021-02-17 with a maximum of 25 deaths. Malawi reached the peak on 2021-01-28 with a maximum of 73 cases. Mozambique reached the peak of daily cases on 2021-01-25 with a maximum of 58 cases. Namibia reached the peak of daily cases on 2021-07-14 with a maximum of 150 cases. South Africa reached the peak of daily cases on 2021-01-06 with a maximum of 844 cases. Zambia reached the peak of daily cases on 2021-07-01 with a maximum of 72 cases. Zimbabwe reached the peak of daily cases on 2021-07-27 with a maximum of 107 cases.
Fig. 1. Daily confirmed cases for the study period.

Fig. 2. Daily death cases for the study period.

Remark 1. We observe that many more deaths have been recorded during the waves of contamination since the start of the pandemic in each country.

B. Daily Cumulative Analytics

Graphics in Fig. 3 show the evolution of the number of contaminations and deaths over the study period in different countries.

Fig. 3. Evolution of the number of confirmed cases and death cases for the study period.
The maximum number of confirmed cases is in South Africa, followed by Zambia, Botswana, Mozambique, Namibia, Zimbabwe, Malawi, Angola. Lesotho has recorded the least number of confirmed cases over the period. As to the number of death cases over the period, South Africa has recorded the greatest number of deaths, followed by Zimbabwe, Zambia, Namibia, Botswana, Malawi, Mozambique, Angola, Lesotho.

Total numbers of confirmed cases and deaths over the study period for each country in the southern African zone are summarized in the Table I.

| Date | Angola | Botswana | Lesotho | Malawi | Mozambique | Namibia | South Africa | Zambia | Zimbabwe |
|------|--------|----------|---------|--------|------------|---------|--------------|--------|----------|
| 2021-09-02 | 48004 | 159317 | 14395 | 60063 | 147066 | 125206 | 2796405 | 206705 | 125118 |

| Date | Angola | Botswana | Lesotho | Malawi | Mozambique | Namibia | South Africa | Zambia | Zimbabwe |
|------|--------|----------|---------|--------|------------|---------|--------------|--------|----------|
| 2021-09-02 | 1235 | 2276 | 403 | 2195 | 1871 | 3394 | 82914 | 3608 | 4449 |

C. Forecast

We use Prophet algorithm for training the model and predicting numbers of infected cases in next three months.

Graphs below show the predicted values of cases for each country where, Black dots refer to the original data, Dark blue line refers to the predicted value (yhat), and Light blue area indicates the prediction interval (yhat-upper and yhat-lower value).

The Fig. 4 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table II sums up forecasts of the total number of confirmed cases for Benin from November 27th, 2021 to December 1st, 2021.

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. \( R^2 = 0.9993, \) \( p \text{ value} \leq 2.2e16 \), degrees of freedom: 588. The Fig. 5. is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

The Fig. 6 below shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table III sums up forecasts of the total number of confirmed cases for Botswana from November 27th, 2021 to December 1st, 2021.
Fig. 5. Plot of actuals versus forecast values.

Fig. 6. COVID-19 Projections using Prophet model. COVID-19 forecast depicts that COVID-19 confirmed cases numbers for Botswana will rise to the tally of 216289.0.

| TABLE III: TOTAL NUMBERS OF EXPECTED CONFIRMED CASES FROM NOVEMBER 27TH, 2021 TO DECEMBER 1ST, 2021 |
|-------------------------------------------------------------------------------------------------|
|                                                                                                  |
| | Total numbers of expected confirmed cases in Botswana |                                                                 |
|-------------------------------------------------------------------------------------------------|
| |Ds       | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|-------------------------------------------------------------------------------------------------|
| | Yhat    | 212821.1   | 213429.6   | 215045.5   | 215662.7   | 216289.0   |
|-------------------------------------------------------------------------------------------------|
| | Yhat_lower | 203341.8   | 204109.8   | 205283.3   | 206337.2   | 206348.6   |
|-------------------------------------------------------------------------------------------------|
| | Yhat_upper | 221641.5   | 222840.6   | 224815.6   | 225109.1   | 225762.0   |

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. $R^2 = 0.9756$, $p \leq 2.2e 16$, degrees of freedom: 588. The Fig. 7. is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

The Fig. 8 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table IV sums up forecasts of the total number of confirmed cases for Lesotho from November 27th, 2021 to December 1st, 2021.

Fig. 7. Plot of actuals versus forecast values.
COVID-19 projections using the Prophet model. COVID-19 forecast depicts that COVID-19 confirmed cases numbers for Lesotho will rise to the tally of 17245.34.

**TABLE IV: Total Numbers of Expected Confirmed Cases from November 27th, 2021 to December 1st, 2021**

| Days       | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|------------|------------|------------|------------|------------|------------|
| Yhat       | 17109.47   | 17141.84   | 17179.07   | 17202.66   | 17245.34   |
| Yhat_lower | 14785.43   | 14712.71   | 14679.37   | 14755.18   | 14736.06   |
| Yhat_upper | 19533.75   | 19581.42   | 19674.63   | 19773.98   | 19792.57   |

Model performance test: the forecast accuracy of the model is represented by R-squared value and p-value of model. \( R^2 = 0.9976, \ p \text{ value} \leq 2.2e16, \) degrees of freedom: 588. The Fig. 9 is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

The Fig. 10 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table V sums up forecasts of the total number of confirmed cases for Malawi from November 27th, 2021 to December 1st, 2021.

**TABLE V: Total Numbers of Expected Confirmed Cases from November 27th, 2021 to December 1st, 2021**

| Days       | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|------------|------------|------------|------------|------------|------------|
| Yhat       | 76454.40   | 76657.19   | 76852.71   | 77077.37   | 77321.63   |
| Yhat_lower | 69113.39   | 68420.99   | 68976.79   | 68768.33   | 69034.14   |
| Yhat_upper | 83024.79   | 83933.19   | 83766.05   | 84152.96   | 84750.72   |
Fig. 10. COVID-19 Projections using Prophet model. COVID-19 forecast depicts that COVID-19 confirmed cases numbers for Malawi will rise to the tally of 77321.63.

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. $R^2 = 0.9837$, $p$ value $\leq 2.2e$ 16, degrees of freedom: 588. The Fig. 11 is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

The Fig. 12 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table VI sums up forecasts of the total number of confirmed cases for Mozambique from November 27th, 2021 to December 1st, 2021.
TABLE VI: Total Numbers of Expected Confirmed Cases from November 27th, 2021 to December 1st, 2021

| Ds       | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|----------|-----------|-----------|-----------|-----------|-----------|
| Yhat     | 193839.9  | 194481.1  | 195087.5  | 195752.0  | 196437.6  |
| Yhat_lower | 180673.6  | 181189.8  | 181316.9  | 181744.2  | 182373.0  |
| Yhat_upper | 208542.7  | 209134.0  | 209247.1  | 210081.2  | 211301.5  |

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. \( R^2 = 0.9807 \), \( p \text{ value} \leq 2.2e16 \), degrees of freedom: 588. The Fig. 13 is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

The Fig. 14 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table VII sums up forecasts of the total number of confirmed cases for Namibia from November 27th, 2021 to December 1st, 2021.

![Fig. 13. Plot of actuals versus forecast values.](image)

![Fig. 14. COVID-19 Projections using Prophet model.](image)

TABLE VII: Total Numbers of Expected Confirmed Cases from November 27th, 2021 to December 1st, 2021

| Ds       | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|----------|-----------|-----------|-----------|-----------|-----------|
| Yhat     | 202768.0  | 203513.7  | 204263.0  | 204948.4  | 205701.4  |
| Yhat_lower | 194252.5  | 195033.6  | 195407.7  | 196215.0  | 197159.5  |
| Yhat_upper | 210456.6  | 211551.0  | 212138.3  | 212836.2  | 214202.9  |

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. \( R^2 = 0.9913 \), \( p \text{ value} \leq 2.2e16 \), degrees of freedom: 588. The Fig. 15 is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

![Fig. 15.](image)
COVID-19 Projections using Prophet model. COVID-19 forecast depicts that COVID-19 confirmed cases numbers for South Africa will rise to the tally of 3763814.

The Fig. 16 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table VIII sums up forecasts of the total number of confirmed cases for South Africa from November 27, 2021 to December 1st, 2021.

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. $R^2 = 0.8893$, $p \text{ value} \leq 2.2e16$, degrees of freedom: 588. The Fig. 17 is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.
The Fig. 18 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table IX sums up forecasts of the total number of confirmed cases for Zambia from November 27, 2021 to December 1st, 2021.

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. \( R^2 = 0.9817, \) \( p \text{ value} \leq 2.2e16, \) degrees of freedom: 588. The Fig. 19 is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

**TABLE IX: TOTAL NUMBERS OF EXPECTED CONFIRMED CASES FROM NOVEMBER 27TH, 2021 TO DECEMBER 1ST, 2021**

|   | Ds     | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|---|--------|------------|------------|------------|------------|------------|
| Yhat | 324979.0 | 326052.9 | 327060.0 | 328216.6 | 329409.4 |
| Yhat_lower | 305665.1 | 306857.7 | 306388.1 | 306612.3 | 308719.7 |
| Yhat_upper | 344534.8 | 345566.2 | 347082.7 | 349226.7 | 350756.9 |

The Fig. 20 shows the prediction period from September 3rd, 2021 to December 1st, 2021 and the Table X sums up forecasts of the total number of confirmed cases for Zimbabwe from November 27th, 2021 to December 1st, 2021.

Model performance test: the forecast accuracy of model is represented by R-squared value and p-value of model. \( R^2 = 0.9817, \) \( p \text{ value} \leq 2.2e16, \) degrees of freedom: 588. The Fig. 21 is the plot of actuals versus forecast values. It shows that predicted values are slightly over and under estimated than actuals but still closed to linear line with high accuracy.

**TABLE X: TOTAL NUMBERS OF EXPECTED CONFIRMED CASES FROM NOVEMBER 27TH, 2021 TO DECEMBER 1ST, 2021**

|   | Ds     | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|---|--------|------------|------------|------------|------------|------------|
| Yhat | 189070.0 | 189760.7 | 190543.7 | 191370.5 | 192198.3 |
| Yhat_lower | 176367.0 | 176755.5 | 178917.5 | 179538.1 | 180466.6 |
| Yhat_upper | 200880.1 | 201588.8 | 203256.3 | 202872.9 | 203884.1 |

**TABLE IX: TOTAL NUMBERS OF EXPECTED CONFIRMED CASES FROM NOVEMBER 27TH, 2021 TO DECEMBER 1ST, 2021**

|   | Ds     | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|---|--------|------------|------------|------------|------------|------------|
| Yhat | 324979.0 | 326052.9 | 327060.0 | 328216.6 | 329409.4 |
| Yhat_lower | 305665.1 | 306857.7 | 306388.1 | 306612.3 | 308719.7 |
| Yhat_upper | 344534.8 | 345566.2 | 347082.7 | 349226.7 | 350756.9 |

**TABLE X: TOTAL NUMBERS OF EXPECTED CONFIRMED CASES FROM NOVEMBER 27TH, 2021 TO DECEMBER 1ST, 2021**

|   | Ds     | 2021-11-27 | 2021-11-28 | 2021-11-29 | 2021-11-30 | 2021-12-01 |
|---|--------|------------|------------|------------|------------|------------|
| Yhat | 189070.0 | 189760.7 | 190543.7 | 191370.5 | 192198.3 |
| Yhat_lower | 176367.0 | 176755.5 | 178917.5 | 179538.1 | 180466.6 |
| Yhat_upper | 200880.1 | 201588.8 | 203256.3 | 202872.9 | 203884.1 |
IV. CONCLUSION

In this paper, we have analyzed different time series COVID-19 data by exploring confirmed and deaths daily reports from January 22, 2020 to September 2nd, 2021 in the southern African countries to begin with. Then, we have used the Prophet model to forecast the total numbers of COVID-19 infected cases from September 3rd, 2021 to December 1st, 2021 by providing graphics and by giving tables containing the total numbers of infected cases from November 27, 2021 to December 1st, 2021 for each country in the zone. Analysis of Fig. 1 and 2 have shown that many more deaths have been recorded during different waves of contamination since the start of pandemic in each country. On the Fig. 3 and the Table I, we have observed that the maximum number of confirmed cases is in South Africa with a total of 2796405 cases, followed by Zambia with 206705 cases, Botswana with 159317 cases, Mozambique with 147066 cases, Namibia with 125206 cases, Zimbabwe with 125118 cases, Malawi with 60663 cases, Angola with 48004 cases. Lesotho has recorded the least number of confirmed cases with 14395 cases over the period. As to the number of death cases over the period, South Africa has recorded more deaths with 82914 deaths, followed by Zimbabwe with 4449 deaths, Zambia with 3608 deaths, Namibia with 3394 deaths, Botswana with 2276 deaths, Malawi with 2195 deaths, Mozambique with 1871 deaths, Angola with 1235 deaths and Lesotho with 403 deaths. Forecasts have depicted that total numbers of confirmed cases will rise to more or less the tally of 62058.50 in Angola according to the Table II, 216289.0 in Botswana according to the Table III, 17245.34 in Lesotho according to the Table IV, 77321.63 in Malawi according to the Table V, 196437.6 in Mozambique according to the Table VI, 205701.4 in Namibia according to the Table VII, 3763814 in South Africa according to the Table VIII, 329409.4 in Zambia according to the Table IX and 192198.3 in Zimbabwe according to the Table X. We hope these information will be helpful to the regulating authority of each country to make impressive arrangements of medical emergency services, quarantine, social distancing etc. to prevent and recover from the pandemic situation raised due to novel COVID-19.
REFERENCES

[1] Bouraima MB, Zonon BI, Qiu Y. The first 100 days: A review and analysis of the evolution of the COVID-19 pandemic in the West African Economic and Monetary Union (WAEMU). Eurasian Journal of Medicine and Investigation. 2020; 4(4): 405-13.

[2] Vishmas BV, Patel A. Hands-on Time series analysis with Python: from basics to bleeding edge techniques. 2020.

[3] Ponce M, Sandhel A: COVID19.Analytics: An R Package to Obtain, Analyze and Visualize Data from Coronavirus Disease Pandemic. arXiv [Preprint] 2021 Available fromhttps://arxiv.org/pdf/2009.01091.pdf

[4] Malik AB. Analysis and Forecast of COVID-19 Pandemic in Pakistan. MedRxiv. [Internet] 2020. Available from: https://www.medrxiv.org/content/10.1101/2020.06.24.20138800v1

[5] Gupta AK, Singh V, Mathur P, Travieso-Gonzalez CM. Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. Journal of Interdisciplinary Mathematics. 2021; 24(1): 89-108.

[6] Milind T, Jain R. COVID-19 Forecast Using Time Series Methods. International Journal of Scientific & Technology Research. 2020;9(8): 45-51.