A Tuned Holt-Winters White-Box Model for COVID-19 Prediction

Seng Hansun, Vincent Charles*, Tatiana Gherman, Subanar, Christiana Rini Indrati

* Corresponding Author

A Tuned Holt-Winters White-Box Model for COVID-19 Prediction

Abstract

The year 2020 has become memorable the moment the novel COVID-19 spread massively around the world to become a pandemic. In this paper, we analyse and predict the future trend of the COVID-19 cases for the top ten countries with the highest number of confirmed cases to date and the top ten countries with the highest growth percentage within the last month. Since many recent works have proposed that the COVID-19 pattern follows an exponential distribution, we use a tuned approach to the Holt-Winters' additive method as a white-box model. Based on the analysis, we found that most of the countries are still presenting an increasing trend of confirmed cases in the near future. Apart from vaccine and drug development, measures such as vigilance, strategic governmental actions, public awareness, and social distancing are unarguably continuously needed to handle the spreading of COVID-19 and avoid the next wave of the outbreak.

Keywords

COVID-19; Future Wave; Holt-Winters Additive Method; Prediction; White-Box Model.

1. Introduction

The year 2020 has become a memorable year for most people around the world. It is the year when a disease called Coronavirus Disease 2019 (COVID-19) has spread massively and took the lives of more than half million people globally within just six months (WHO, 2020). Although the disease could be traced back to the late of 2019, when it was believed that Wuhan City in the Hubei Province in China had become the epicentre of the disease (Li *et al*., 2019; Abd El-Aziz & Stockand, 2020; Acter *et al*., 2020), it is in 2020 that the disease became a pandemic (Spinelli & Pellino, 2020).

COVID-19 is a major public health issue that is caused by a novel type of coronavirus. As announced by the International Committee on Taxonomy of Viruses, the virus is called *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-CoV-2) (Lai *et al*., 2020). This virus is known to have great commonalities with two previous pandemics, which took place in 2002 and 2012, respectively; these are the Severe Acute Respiratory Syndrome (SARS) coronavirus and the Middle Eastern Respiratory Syndrome (MERS) coronavirus (Park *et al*., 2020). It further counts with a high reproductive number, a long period of incubation, and a low rate of case fatality compared to SARS and MERS (Xie & Chen, 2020).

Due to its importance and impacts on society, a growing number of research related to COVID-19 can be seen daily. Some focus on drug and vaccine creation (Dhama *et al*., 2020; Liu *et al*., 2020), some on the strategic management of the virus (Block *et al*., 2020; Mandal *et al*., 2020), and some others on the prediction of the growth and trend of the virus. Research on the prediction of COVID-19 has gained a special interest because it could help the Government, the policymakers, other stakeholders, and the society at large, to take precautionary actions when facing this crisis. Various types of methods have been used for prediction purposes, such as mathematical models, time series analyses, and even sophisticated soft computing algorithms.

Fanelli and Piazza (2020) used a simple susceptible-infected-recovered-deaths (SIRD) model to analyse and forecast the COVID-19 spreading in China, Italy, and France. Meanwhile, Wangping *et al.* (2020) developed an infectious disease dynamic extended susceptible-infected-removed (eSIR) model to estimate the COVID-19 trend in Italy. Singhal *et al.* (2020) developed a Gaussian mixture model to predict the COVID-19 in several countries, such as India, Italy, and the United States of America (USA). Torrealba-Rodriguez *et al*. (2020) used both Gompertz and Logistic models, as well as Artificial Neural Network, to predict COVID-19 confirmed cases in Mexico. Similarly, Yang *et al.* (2020) also combined the susceptible-exposed-infection-removed (SEIR) model and an Artificial Intelligence (AI) approach to predict the epidemic. As for the time series analysis and modelling, Petropoulos and Makridakis (2020) applied a univariate time series model from an exponential smoothing family. Qi *et al*. (2020) experimented on 14-day exponential moving averages to find the association between COVID-19 with temperature and relative humidity in 30 provinces in China. Lastly, Cao *et al*. (2020) also implemented different kinds of methods for epidemic prediction of COVID-19, including the exponential smoothing method, the autoregressive integrated moving average (ARIMA), and the ARIMAX method.

Moreover, many researchers have used powerful soft computing methods in COVID-19 prediction and analysis. Swapnarekha *et al.* (2020) examined that machine learning methods have been efficiently used in numerous literatures for prediction and diagnosis of COVID-19. Chimmula and Zhang (2020) predicted the COVID-19 transmission in Canada by using the long short-term memory (LSTM) network. Kırbaş *et al.* (2020) also used and compared the LSTM method with ARIMA and nonlinear autoregression neural network (NARNN) in forecasting the COVID-19 cases. Mei *et al*. (2020) implemented various kind of AI methods, such as convolutional neural network (CNN), support vector machine (SVM), random forest (RF), and multilayer perceptron (MLP) during their study to find an AI-enabled system that could rapidly diagnose COVID-19 suspects. Yan *et al.* (2020) applied the supervised XGBoost classifier as the predictor model in predicting COVID-19 patients at the highest risk to reduce the mortality rate. Lastly, Hasan (2020) proposed and reported the effectiveness of a hybrid model that uses ensemble empirical mode decomposition (EEMD) and artificial neural network (ANN) in COVID-19 prediction.

Although prediction models for COVID-19 have multiplied in number, some critical appraisals have emerged in the academic literature. Wynants *et al.* (2020) specifically screened around 4900 titles and further examined 51 studies that describe 66 prediction models for COVID-19. They found that the proposed models are poorly reported, give a high risk of bias, with overly optimistic results. Furthermore, Roda *et al.* (2020) argued that predictions by using sophisticated models, also known as black-box models (such as Machine Learning methods), might not be more reliable than using a simpler one. This is because there is still much we do not know or understand yet about the pandemic; hence, we need models that we can clearly explain to stakeholders how they behave, how they produce predictions, and what the influencing variables might be; in other words, we need white-box models. This is the rationale that guided us in the choice of model for the present paper. The following Table 1 provides a brief comparison between black-box and white-box models, highlighting their advantages and disadvantages.

Table 1.
A Comparison Between Black-Box Models and White-Box Models.

Black-box models	White-box models
Address highly non-linear structures.	Address linear or stepwise linear or curve linear structures.
Logically well-defined and mathematically complex.	Logically and mathematically well defined, and simple.
Large number of parameters; hence, models are high-dimensional.	Small number of parameters.
Large number of features.	Relatively small number of features.
Statistical hypothesis testing is irrelevant.	Statistical hypothesis testing is relevant.
High computational complexity.	Low computational complexity.
Lack clarity around inner workings.	The input-output relationship is visible, and the process through which the output is produced is also visible.
Large data set.	Relatively small data set.
Do not warrant any statistical distributional assumptions.	Warrant statistical distributional assumptions.
Modelling is usually a trial and error and iterative process.	Modelling is less of a trial and error process and more of a systematic approach.
Guided by rules of thumb.	Guided by established criteria.
Results depend on the hyperparameter tuning strategy.	Results depend on the statistical estimation properties.
Lower explainability or interpretability.	Higher explainability or interpretability.
Lower transparency and accountability.	Higher transparency and accountability.
Higher accuracy.	Relatively lower accuracy.

In this study, we aim to analyse and predict the future trend of the COVID-19 cases for the top ten countries with the highest number of confirmed cases to date (14 June 2020) and the top ten countries with the highest growth percentage within the last month. After China lifted Wuhan City lockdown on 8 April 2020 (BBC, 2020), a growing number of countries have also lifted or eased the lockdown restrictions in May and June 2020 (Pharmaceutical Technology, 2020). Therefore, the prediction of COVID-19 during this time period could reveal which countries may be affected mostly by the next wave of COVID-19. Since many recent works imply that the COVID-19 pattern follows an exponential distribution (Leung *et al.*, 2020; Tuli *et al.*, 2020), we will use a modified approach to Holt-Winters' additive method, as reported in Hansun *et al.* (2019).

2. Data Source and Applied Algorithm

Since the early time of the COVID-19 outbreak, a team of scientists at Johns Hopkins University (JHU) have responded to the public health crisis by developing an online real-time interactive dashboard to visualise and track reported cases of COVID-19 (Dong *et al.*, 2020). They used GitHub as a data repository for COVID-19 Dashboard, which is operated by the JHU Center for Systems Science and Engineering (JHU CSSE, 2020). The repository consists of nation-wide and global data, such as confirmed, recovered, and death cases (Tiwari *et al.*, 2020). In this study, we use the confirmed cases around the world from 22 January 2020 to 14 June 2020 to predict the future values for the top ten countries with the highest number of confirmed cases to date (14 June 2020) and the top ten countries with the highest growth percentage within the last month.

In this study, a modified approach to the Holt-Winters' additive method is applied. Holt-Winters (HW) is one of the most commonly used methods in the exponential smoothing family (Hansun, 2017). In 2019, Hansun *et al.* (2019) proposed new formulas for finding the initial values in the HW additive method. In summary, the proposed formulas for the initial overall smoothing (S) and the initial trend smoothing (b) are shown in Eq. (1) and Eq. (2) below:

$$S = \frac{Ly_L + (L-1)y_{(L-1)} + \cdots + (L-n+2)y_{(L-n+2)} + (L-n+1)y_{(L-n+1)}}{L + (L-1) + \cdots + (L-n+2) + (L-n+1)}, \tag{1}$$

$$b = \frac{1}{L^2}\left(\frac{2Ly_{2L} + (2L-1)y_{(2L-1)} + \cdots + (L+2)y_{(L+2)} + (L+1)y_{(L+1)}}{2L + (2L-1) + \cdots + (L+2) + (L+1)} - \frac{Ly_L + (L-1)y_{(L-1)} + \cdots + 2y_2 + y_1}{L + (L-1) + \cdots + 2 + 1}\right), \tag{2}$$

where L is the season length, n is the span period of the forecasting formula, and y is the real data. Meanwhile, the initial values for the seasonal indices are computed by averaging every observed season, dividing every observed value by the season's average, and lastly, by averaging each of these numbers across the observed seasons.

By using all the initial values, we follow the HW additive method procedure to find the overall smoothing, trend smoothing, and seasonal smoothing, as shown in Eq. (3) – Eq. (5):

$$S_t = \alpha(y_t - I_{t-L}) + (1-\alpha)(S_{t-1} + b_{t-1}), \tag{3}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, \tag{4}$$

$$I_t = \beta(y_t - S_{t-1} - b_{t-1}) + (1-\beta)I_{t-L}, \tag{5}$$

where S_t is the overall smoothing, b_t is the trend smoothing, I_t is the seasonal smoothing, y_t is the actual data, and t is an index denoting a time period. The season length (L) refers to the number of data points that indicate the beginning of a new season. α, β, and γ are constant parameters between 0 and 1, which should be estimated in such a way so as to minimise the error measurements. Lastly, the forecasted value can be calculated using Eq. (6):

$$F_{t+m} = S_t + mb_t + I_{t-L+m}, \tag{6}$$

where F_{t+m} is the forecast at m periods ahead.

The main difference between the modified approach and the original HW additive method lies in the process of determining the initial values for overall smoothing and trend smoothing. The initial value for overall smoothing in the original approach can be found simply by using the last observed value. Meanwhile, in the modified approach, the weighting factor calculation from several recent observations is applied to get the initial value for overall smoothing, as shown in Eq. (1). Similarly, while the initial value for trend smoothing in the original approach only concerns the difference between observations in the first and second seasons, in the modified approach, we added the weighting factor calculation from those seasons before taking the difference, as shown in Eq. (2). Moreover, we also implement a brute force approach to get the best constant parameters (α, β, and γ) in the training phase of this study.

3. Data Analysis

In the present work, we use the global confirmed cases of COVID-19 from JHU CSSE, which have been recorded daily from 22 January 2020 to 14 June 2020. However, as previously mentioned, in the data analysis, we are not only using the highest number of confirmed cases, but also the highest growth percentage of confirmed cases in the world within the last month. The motivation behind this decision is to predict the changing situation of COVID-19 more accurately, especially due to the easing of the lockdown restrictions that happened in most countries in May and June 2020. Therefore, there are two kinds of data that are of our interest, *i.e.*, the data concerning the top ten countries with the highest number of confirmed cases to date and the data concerning the top ten countries with the highest growth percentage of confirmed cases within the last month. Moreover, to help with the data analysis, we use Python 3.7 programming language within the Anaconda 3 environment. The following Figure 1 shows the process that we followed in this paper.

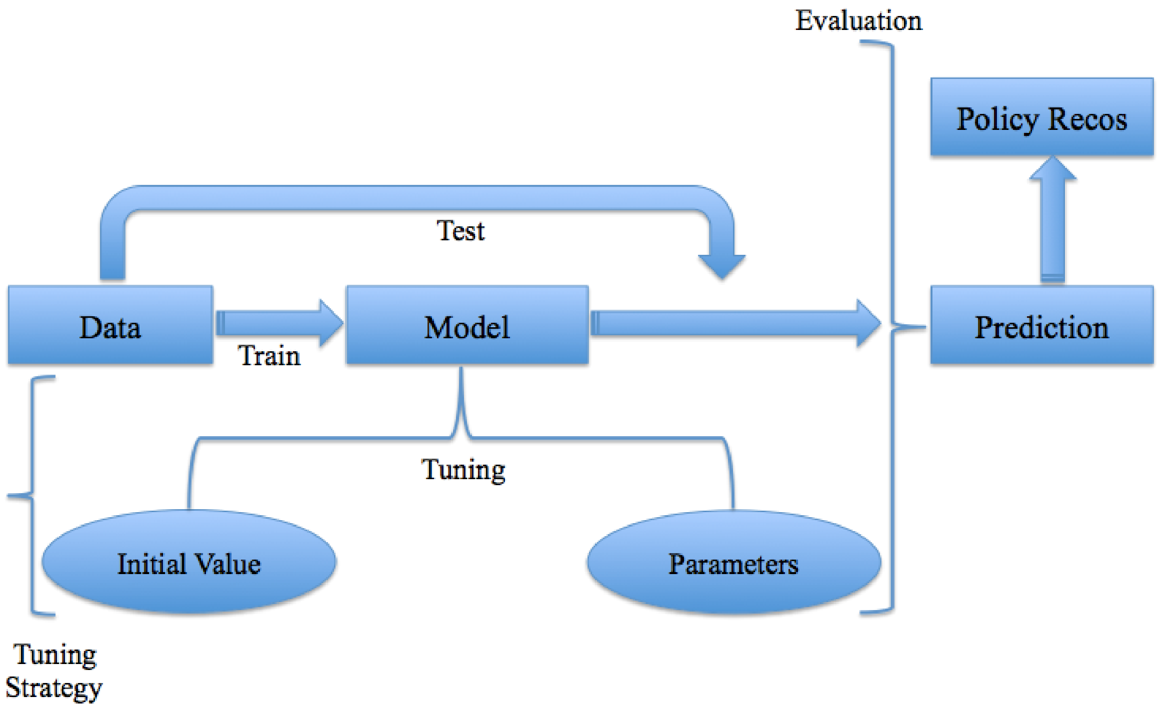


Fig. 1. Process Diagram.

First, the data set is divided into two parts, one for the training phase and another one for the testing phase. The tuning strategy is implemented in the training phase, where the initial values (overall smoothing, trend smoothing, seasonal indices) and parameters (α, β, γ) are chosen to give the smallest error from the model implemented. Next, the best-found parameters are used in the testing phase of the model to get the prediction of future numbers of COVID-19 confirmed cases. Then, the prediction results can be used as a foundation for the policy recommendations of the countries considered in this study.

3.1 Top Ten Countries with the Highest Growth Percentage of Confirmed Cases

Figure 2 shows the general algorithm to import all the needed libraries, read the CSV data, get the countries' names, and extract the start date up to the last date of the data given.

```
# import all needed libraries
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# read csv data
data = pd.read_csv("time_series_covid19_confirmed_global.csv", index_col=0)

# get the countries' name
countries = data["Country/Region"]

# extract the starting date up to the last date
x = list(data.columns)
x = x[3:]
```

Fig. 2. General starting procedure.

Next, as shown in Figure 3, we performed the analysis of the percentage growth by getting the last four weeks' dates from the data set. Then, we extracted the data records for the associated dates and calculated the growth percentage of confirmed COVID-19 cases. We saved the growth percentage data in a new data column and got the rank of each country. Lastly, we selected the top ten rank and plotted the result, as depicted in Figure 4.

```
## percentage growth analysis ##
# get the last four week dates
xt = [x[-i] for i in range(1, 30, 7)]

# extract the data records for the associated week dates
yt, yt1, yt2, yt3, yt4 = ([] for i in range(5))

for i in range(len(countries)):
    yt.append(data.iloc[i][xt[0]])
    yt1.append(data.iloc[i][xt[1]])
    yt2.append(data.iloc[i][xt[2]])
    yt3.append(data.iloc[i][xt[3]])
    yt4.append(data.iloc[i][xt[4]])

# calculate the growth percentage based on four last weeks data
dt, dt_1w, dt_2w, dt_3w, dt_4w = ([] for i in range(5))

for i in range(len(countries)):
    dt_1w.append((yt[i] - yt1[i]) / yt1[i] if yt1[i] else 1)
    dt_2w.append((yt1[i] - yt2[i]) / yt2[i] if yt2[i] else 1)
    dt_3w.append((yt2[i] - yt3[i]) / yt3[i] if yt3[i] else 1)
    dt_4w.append((yt3[i] - yt4[i]) / yt4[i] if yt4[i] else 1)
    dt.append((dt_1w[i] + dt_2w[i] + dt_3w[i] + dt_4w[i]) / 4)

# save the growth percentage in a new data column 'Percentage'
data["Percentage"] = dt

# get the rank of each country and save it in a new data column 'Rank'
data["Rank"] = data["Percentage"].rank(ascending=False)
```

```
# select top ten rank and save it under new data column 'Selected'
data["Selected"] = data["Rank"] <= 10

# plot top ten countries with the highest growth percentage within a month
for i in range(len(countries)):
    if data["Selected"][i]:
        y = data.iloc[i,3:-3]
        if np.isnan(data.index[i]):
            lbl = str(data.iloc[i,0])
        else:
            lbl = str(data.iloc[i,0]) + " - " + str(data.index[i])
        plt.plot(x, y, label=lbl)

ind = [i for i in range(0, len(x), 7)]
date = [x[i] for i in ind]
plt.xticks(ind, date, rotation=60)

plt.xlabel("Time")
plt.ylabel("Number")
plt.title("COVID-19 Confirmed - Growth Percentage")
plt.legend()
plt.savefig('10 Growth Percentage.jpg', bbox_inches = 'tight')
plt.figure(figsize=(20,10))
```

Fig. 3. Analysis of the Growth Percentage of Confirmed COVID-19 Cases.

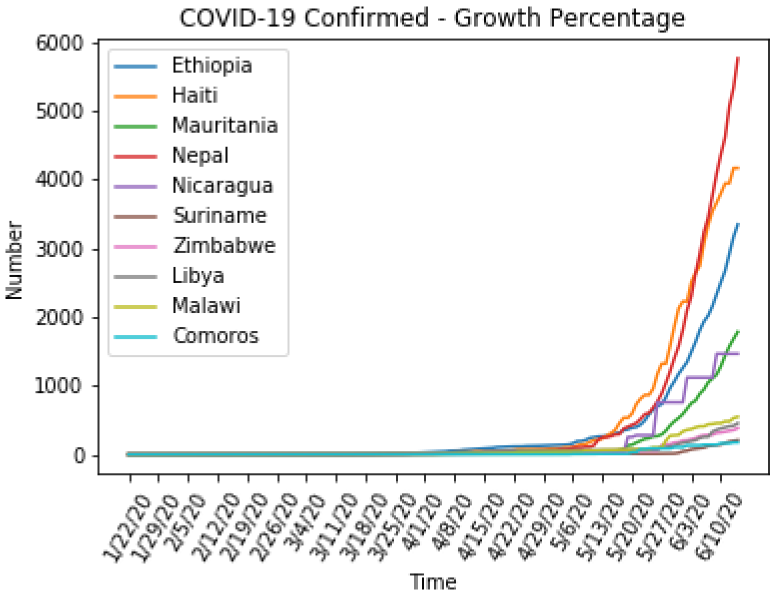


Fig. 4. Top Ten Countries with the Highest Growth Percentage of COVID-19 Cases.

As shown in Figure 4, most of the top ten countries with the highest growth percentage within the last month (17 May – 14 June 2020) are located in Africa. This result aligns with the global comparison report from Ritchie *et al*. (2020) and Roser *et al*. (2020), but may change due to the rapidly changing situation of COVID-19 worldwide.

3.2 Top Ten Countries with the Highest Number of Confirmed COVID-19 Cases

To get the top ten countries with the highest number of confirmed COVID-19 cases, we need to get the rank of each country based on the number of the last date confirmed cases, save it in a new data column, select only the top ten rank, and plot the result (Figure 5). The top ten countries with the highest number of confirmed COVID-19 cases (until 14 June 2020) is shown in Figure 6.

```
## up-to-date confirmed cases analysis ##
# get the rank of each country based on the number of last date confirmed cases,
# save it in a new data column 'Rank_Global'
# select top ten rank and save it under new data column 'Selected_Global'
data["Rank_Global"] = data[x[-1]].rank(ascending=False)
data["Selected_Global"] = data["Rank_Global"] <= 10

# plot top ten countries based on the number of last date confirmed cases
for i in range(len(countries)):
    if data["Selected_Global"][i]:
        y = data.iloc[i,3:-5]
        if np.isnan(data.index[i]):
            lbl = str(data.iloc[i,0])
        else:
            lbl = str(data.iloc[i,0]) + " - " + str(data.index[i])
        plt.plot(x, y, label=lbl)

ind = [i for i in range(0, len(x), 7)]
date = [x[i] for i in ind]
plt.xticks(ind, date, rotation=60)

plt.xlabel("Time")
plt.ylabel("Number")
plt.title("COVID-19 Confirmed - Global")
plt.legend()
plt.savefig('10 Confirmed Global.jpg', bbox_inches = 'tight')
plt.figure(figsize=(20,10))
```

Fig. 5. Analysis of the Highest Number of Confirmed COVID-19 Cases.

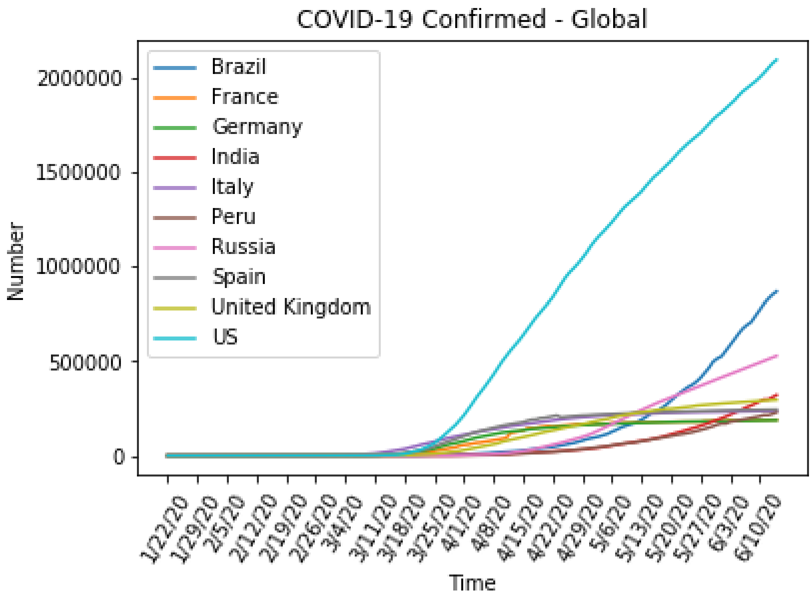


Fig. 6. Top Ten Countries with the Highest Number of Confirmed COVID-19 Cases.

The interested reader can access the codes illustrated in Sections 3.1 and 3.2 above in the GitHub repository at https://github.com/profvc/Prediction_COVID19.git.

4. Prediction Results

In this section, we show the prediction results for the top ten countries with the highest number of confirmed COVID-19 cases to date and the top ten countries with the highest growth percentage of cases within the last month by using a modified approach to the HW additive method, as explained in Section 2. We divide the data set into two parts for the training and testing phases, with 80:20 ratio. For the training phase, we use the data from 22 January 2020 to 16 May 2020 (116 data points), while for the testing phase, we use the data from 17 May 2020 to 14 June 2020 (29 data points). From the training phase, we get the best parameters for α, β, and γ, together with their Mean Absolute Percentage Error (MAPE) values for each country, as shown in Table 2 and Table 3. In this experiment, we use MAPE to estimate the prediction error for each country. We also use seven seasonal lengths with four span data periods as the input parameters for the modified approach.

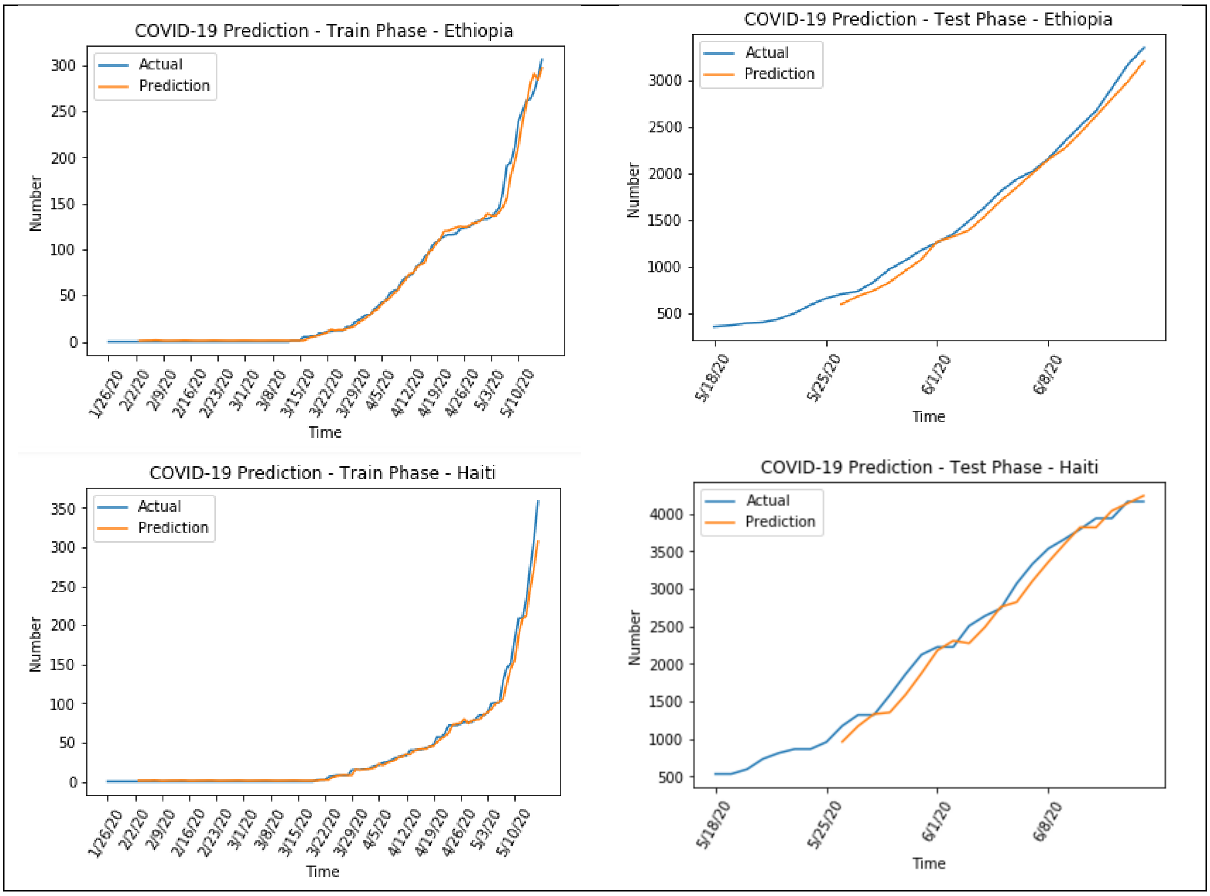
Table 2.
Best Parameters for the Top Ten Countries with the Highest Growth Percentage of Confirmed COVID-19 Cases.

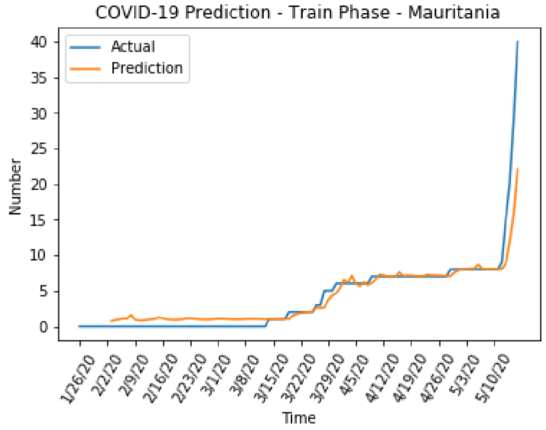
Country Name	Best α	Best β	Best γ	MAPE
Ethiopia	0.46	0.44	0.20	7.8364
Haiti	0.86	0.30	0.04	8.4836
Mauritania	0.46	0.58	0.02	8.6591
Nepal	0.98	0.04	0.04	8.2386
Nicaragua	0.88	0.32	0.02	6.2598
Suriname	0.44	0.56	0.00	6.3195
Zimbabwe	0.92	0.26	0.02	7.8247
Libya	0.98	0.22	0.00	6.3946
Malawi	0.96	0.28	0.00	7.2290
Comoros	0.52	0.36	0.04	6.1156

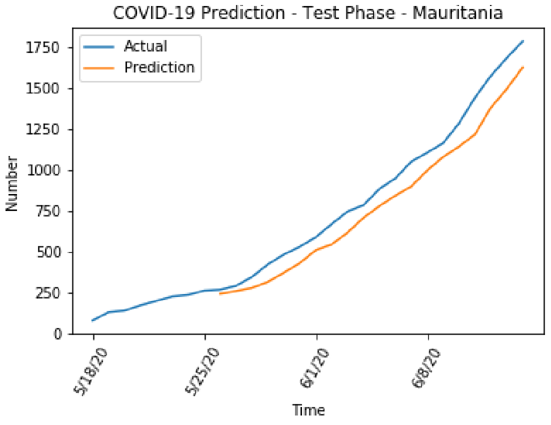
Table 3.

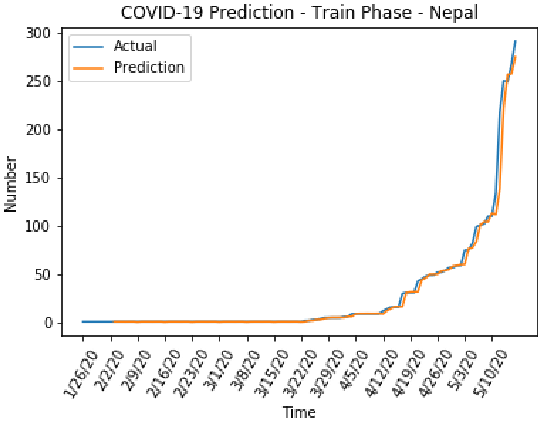
Best Parameters for the Top Ten Countries with the Highest Number of Confirmed COVID-19 Cases

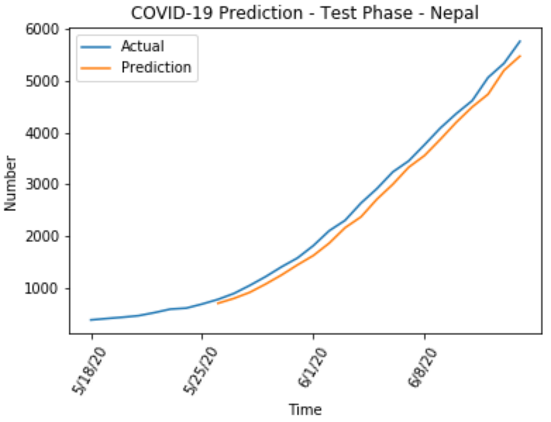
Country Name	Best α	Best β	Best γ	MAPE
Brazil	0.98	0.24	0.28	11.5175
France	0.94	0.00	0.62	5.8901
Germany	0.90	0.04	0.68	4.5761
India	0.90	0.14	0.32	6.6059
Italy	0.80	0.04	0.98	4.9094
Peru	0.98	0.36	0.12	11.8567
Russia	0.90	0.32	0.44	8.5482
Spain	0.98	0.08	0.52	9.3054
United Kingdom	0.72	0.04	0.64	6.5180
US	0.92	0.00	0.98	4.5649

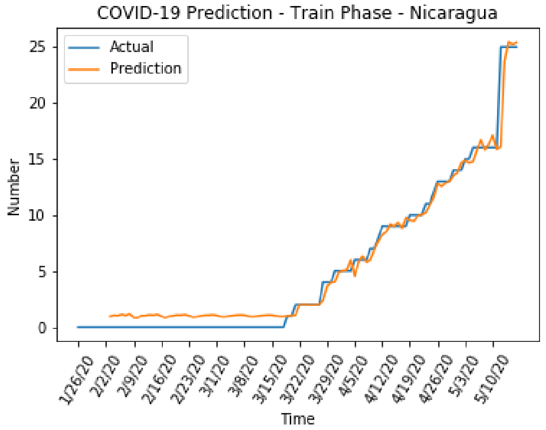


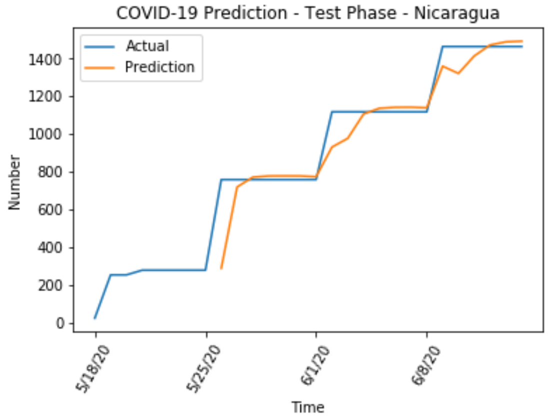

COVID-19 Prediction - Train Phase - Mauritania
Actual
Prediction
Number
Time

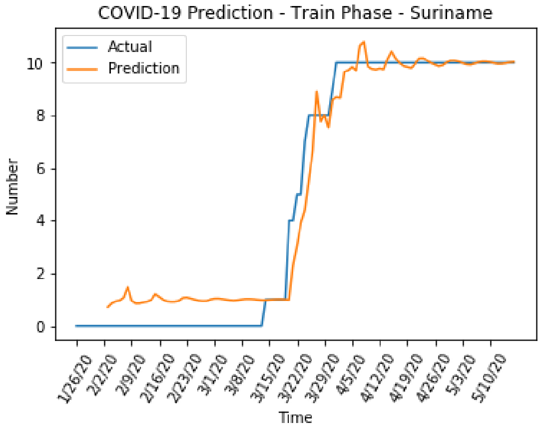

COVID-19 Prediction - Test Phase - Mauritania
Actual
Prediction
Number
Time

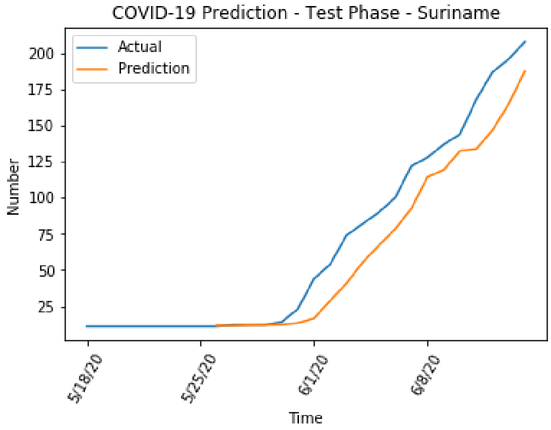

COVID-19 Prediction - Train Phase - Nepal
Actual
Prediction
Number
Time


COVID-19 Prediction - Test Phase - Nepal
Actual
Prediction
Number
Time


COVID-19 Prediction - Train Phase - Nicaragua
Actual
Prediction
Number
Time


COVID-19 Prediction - Test Phase - Nicaragua
Actual
Prediction
Number
Time


COVID-19 Prediction - Train Phase - Suriname
Actual
Prediction
Number
Time


COVID-19 Prediction - Test Phase - Suriname
Actual
Prediction
Number
Time

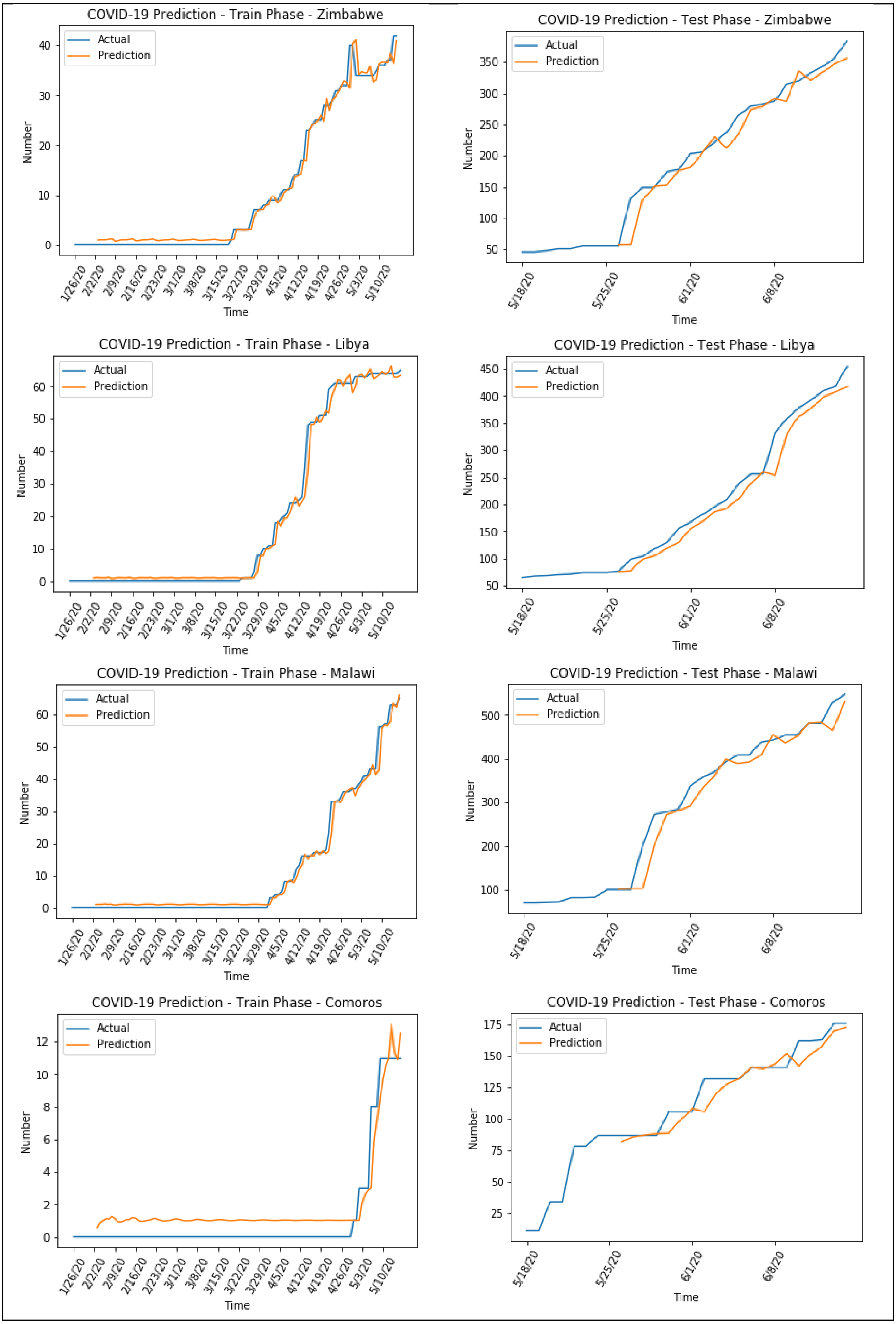
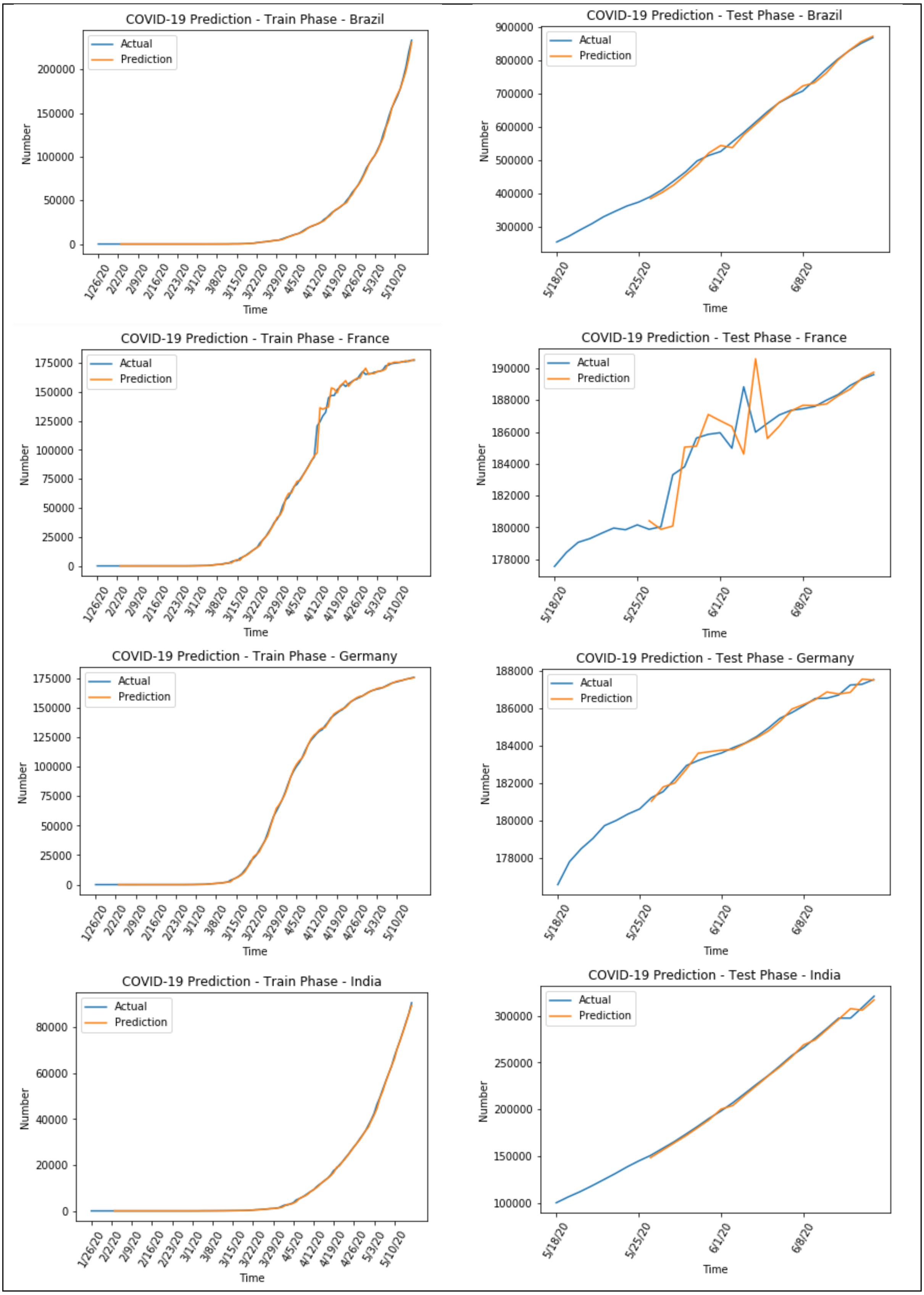
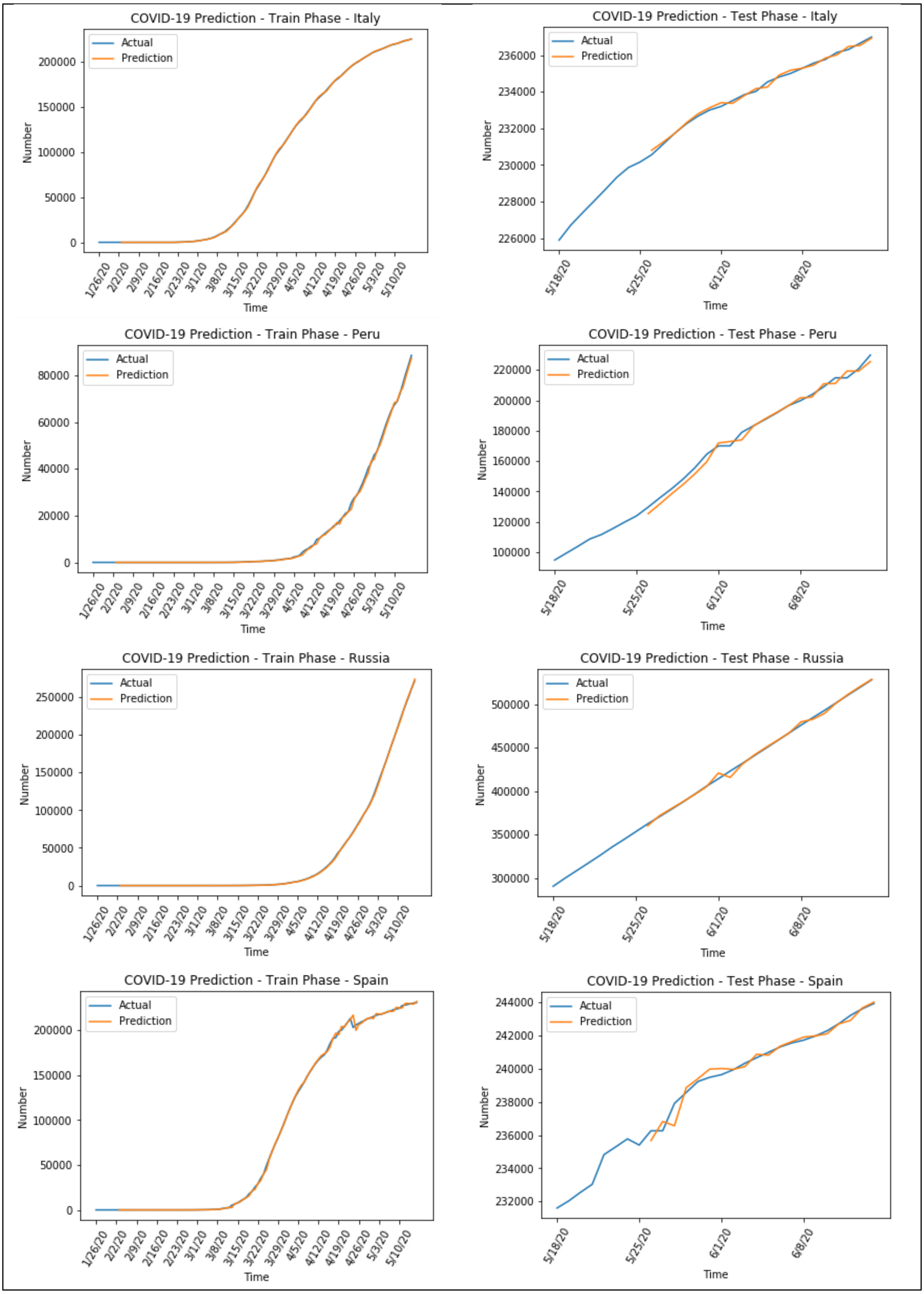


Fig. 7. Prediction Results for the Top Ten Countries with the Highest Growth Percentage of COVID-19 Cases.


COVID-19 Prediction - Train Phase - Brazil
Actual
Prediction
Number
Time
COVID-19 Prediction - Test Phase - Brazil
Actual
Prediction
Number
Time
COVID-19 Prediction - Train Phase - France
Actual
Prediction
Number
Time
COVID-19 Prediction - Test Phase - France
Actual
Prediction
Number
Time
COVID-19 Prediction - Train Phase - Germany
Actual
Prediction
Number
Time
COVID-19 Prediction - Test Phase - Germany
Actual
Prediction
Number
Time
COVID-19 Prediction - Train Phase - India
Actual
Prediction
Number
Time
COVID-19 Prediction - Test Phase - India
Actual
Prediction
Number
Time


COVID-19 Prediction - Train Phase - Italy
COVID-19 Prediction - Test Phase - Italy
COVID-19 Prediction - Train Phase - Peru
COVID-19 Prediction - Test Phase - Peru
COVID-19 Prediction - Train Phase - Russia
COVID-19 Prediction - Test Phase - Russia
COVID-19 Prediction - Train Phase - Spain
COVID-19 Prediction - Test Phase - Spain
Actual
Prediction
Number
Time

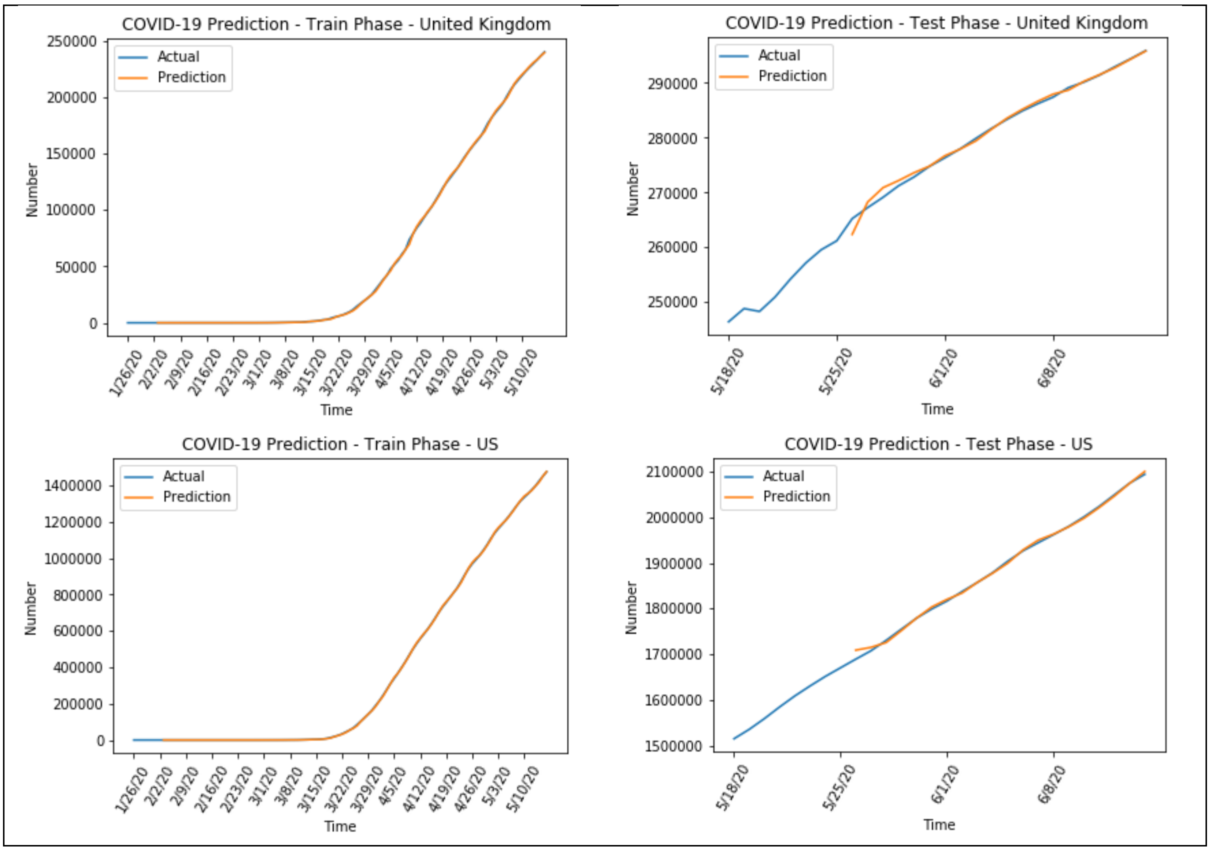


Fig. 8. Prediction Results for the Top Ten Countries with the Highest Number of Confirmed COVID-19 Cases.

Figure 7 and Figure 8 visually exhibit the prediction results for both sets of data in both the training and testing phases. Using the best parameters found as shown in Table 2 and Table 3, we predict the future values for each country in the testing phase. Table 4 shows the MAPE and future prediction of confirmed COVID-19 cases for the top ten countries with the highest growth percentage of COVID-19 cases within the last month. Table 5 shows the values of the same attribute for the top ten countries with the highest number of confirmed COVID-19 cases. Moreover, we also show the upward or downward trend percentage for the prediction that is compared to the last known values in the data set.

Table 4.

Future Trend Prediction for the Top Ten Countries with the Highest Growth Percentage of Confirmed COVID-19 Cases.

Country Name	MAPE	Future Prediction	Trend
Ethiopia	5.8079	3 459	Up + 3.41%
Haiti	6.1507	4 225	Up + 1.44%
Mauritania	13.9907	1 711	Down – 4.07%
Nepal	7.6191	5 923	Up + 2.82%
Nicaragua	7.0589	1 491	Up + 1.83%
Suriname	21.1618	197	Down – 5.46%
Zimbabwe	7.8912	390	Up + 1.78%
Libya	8.1164	469	Up + 3.26%
Malawi	7.3473	554	Up + 1.25%
Comoros	5.2031	177	Up + 0.54%

Table 5.
Future Trend Prediction for the Top Ten Countries with the Highest Number of Confirmed COVID-19 Cases.

Country Name	MAPE	Future Prediction	Trend
Brazil	1.4276	894 678	Up + 3.12%
France	0.5540	189 959	Up + 0.19%
Germany	0.0958	187 774	Up + 0.14%
India	0.9635	330 842	Up + 3.09%
Italy	0.0519	237 372	Up + 0.16%
Peru	1.5836	234 366	Up + 2.02%
Russia	0.3835	539 696	Up + 2.16%
Spain	0.1120	244 354	Up + 0.18%
United Kingdom	0.1997	297 510	Up + 0.55%
US	0.2092	2 114 721	Up + 0.99%

From Table 4, we can conclude that almost all of the ten countries still have an increasing trend of COVID-19 confirmed cases, except for Suriname and Mauritania, who present a decreasing trend instead. Among the top ten countries with the highest growth percentage of confirmed COVID-19 cases within the last month, Ethiopia, Libya, and Nepal are the three countries that present more than 2.5% increasing trend percentage, so they are most likely to have an increasing number of confirmed COVID-19 cases in the near future.

Moreover, if we check the prediction for the top ten countries with the highest number of confirmed COVID-19 cases, we see that all these countries still have an increasing trend in the near future. However, most of these countries have a relatively small increasing trend percentage, which is less than 1%. The exceptions are Brazil, India, Russia, and Peru; these are countries that have more than 2% increasing trend percentage of confirmed cases. The following Figure 9 and Figure 10 visually depict the trend percentage based on the last column in Table 4 and Table 5, respectively.

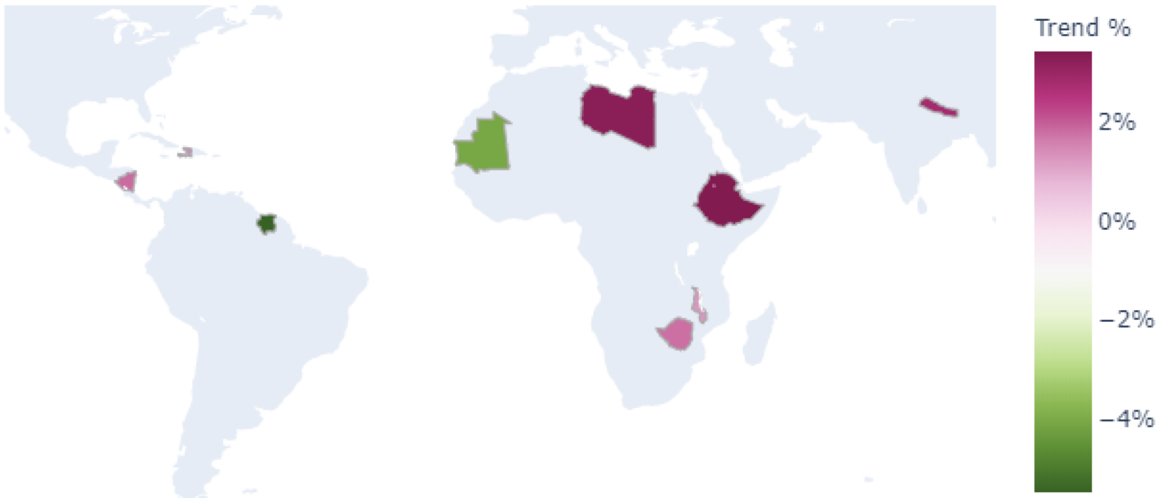


Fig. 9. Trend Percentage Projection for the Top Ten Countries with the Highest Growth Percentage of Confirmed COVID-19 Cases.

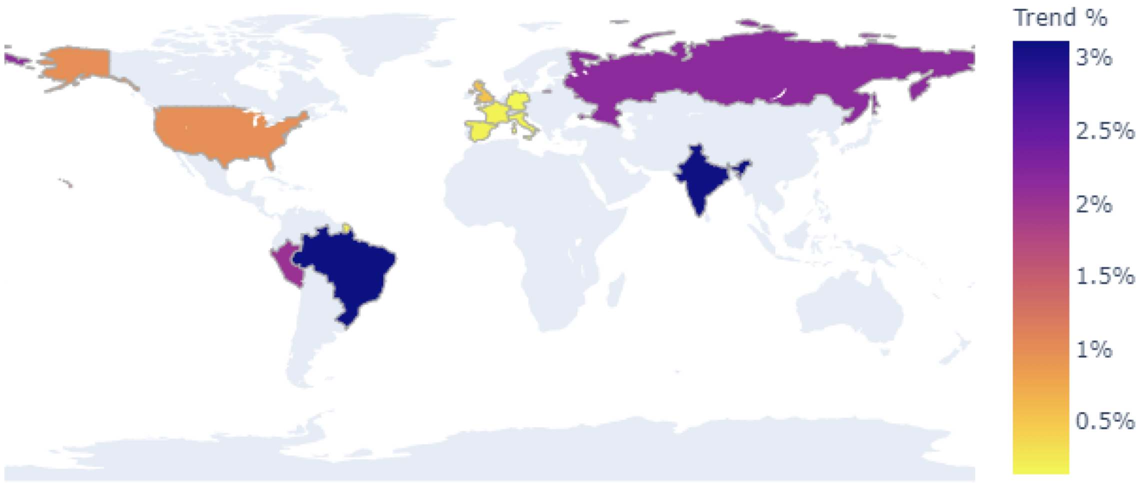


Fig. 10. Trend Percentage Projection for the Top Ten Countries with the Highest Number of Confirmed COVID-19 Cases.

5. Implications for practice and conclusions

As a response to the changing global situation caused by the COVID-19 pandemic, we analysed and predicted the future trend for the top ten countries with the highest number of confirmed COVID-19 cases to date (14 June 2020) and the top ten countries with the highest growth percentage within the last month (May-June 2020). We used a tuned approach to the Holt Winters' additive method as a white-box model and found that the applied method could predict the future trend of the most affected countries. In this sense, results indicate that most of the top ten countries with the highest growth percentage are located in the African continent. Moreover, the top ten countries with the highest number of confirmed COVID-19 cases are countries that have a high population number.

Among the top ten countries with the highest growth percentage, Suriname and Mauritania are the two countries that have a decreasing trend prediction. On the other hand, three countries have more than 2.5% increasing trend percentage; these countries are Ethiopia (3.41%), Libya (3.26%), and Nepal (2.82%). Moreover, all countries included in the top ten countries with the highest number of confirmed COVID-19 cases have an increasing trend prediction, although for most of them, the percentage is relatively small, less than 1%. The exceptions are posed by Brazil (3.12%), India (3.09%), Russia (2.16%), and Peru (2.02%).

A debate has recently ensued with regards to whether the COVID-19 pandemic is unfolding in "waves" or it is only "one big wave" with ups and downs, although the World Health Organization stated that there is no evidence of seasonal variations (The Guardian, 2020). Nevertheless, independent of how we wish to perceive the behaviour of the virus, our results support the argument that measures such as vigilance and strategic governmental actions, effective communication strategies, public awareness and education, and social distancing are unarguably continuously needed to handle the spreading of COVID-19, especially considering that vaccine and drug development will take a long time (about 12-18 months, according to 'The race against COVID-19', 2020) and herd immunity has not yet been achieved (Graeden *et al*., 2020). As Sabat *et al.* (2020) stated, "policymakers and public health experts have to persuade their citizens to make behavior changes and respect future containment interventions while facing the difficulty of enforcing such regulations" (p. 2). Otherwise, the near future will inevitably bring the “next wave” or the “next up of the one big wave” of the COVID-19 outbreak.

There is, however, the critical issue of striking a balance between saving lives and saving livelihoods (Sabat *et al*., 2020), with a large part of the population becoming more and more concerned about the negative impact of the pandemic on the economy. The proposed model and our results could be used as an early warning system by means of helping relevant governments and stakeholders to monitor the current situation and use our forecasts to prevent further transmissions. The lifting of lockdown restrictions has become a subject of increasing public debate and scrutiny, with critical questions being raised in regard to timing, risks, and consequences (The New York Times, 2020). Guided by the short-term forecasts, the localised lockdown periods and other measures could be adjusted accordingly.

Large amounts of data within the framework of big data are constantly being generated at an exponential rate; and by depicting more and more complex scenarios and interrelationships, these data hold the potential to expand the opportunities for making better decisions at strategic, tactical, and operational levels (Charles, Tavana, & Gherman, 2015). Transforming these data into meaningful knowledge (Charles & Gherman, 2013) in a timely manner represents a key factor of success in the fight against the COVID-19 pandemic. As we noted in the introductory section, recent studies have used a range of artificial intelligence and machine learning approaches (or black-box models) to develop sophisticated prediction models to make sense of these rather large data sets; but, as Roda *et al.* (2020) argued, these models might not be more reliable than using a simpler one. This is because there is still much we do not know or understand yet about the pandemic; hence, we need models that we can clearly explain to stakeholders how they behave, how they produce predictions, and what the influencing variables might be.

In view of this, using a white-box model, such as Hansun *et al.* (2019)’s tuned HW method, was considered to be suitable. Of course, there is always a trade-off between using black-box and white-box models. While black-box models provide higher accuracy and are said to outperform white-box models even when dealing with less complex scenarios, they lack clarity around inner workings; white-box models avoid such shortcomings by allowing higher explainability or interpretability. In our case, having higher interpretability is meaningful, since as mentioned, the COVID-19 pandemic is still less than well understood; hence, the fact that we do not yet count with enough practical experience deems the ability to interpret the models we create as rather valuable. Furthermore, future research could complement these analyses with additional insights gathered from more qualitative approaches (Charles & Gherman, 2018), which would help to provide a more comprehensive picture of the unstructured data gathered, and thus, of the COVID-19 pandemic and the measures taken to fight against it; this, in turn, would generate even more societal value.

Acknowledgments

The authors would like to thank the Editor and the two anonymous reviewers for their valuable feedback on the previous version of this manuscript.

References

Abd El-Aziz, T. M., & Stockand, J. D. (2020). Recent progress and challenges in drug development against COVID-19 coronavirus (SARS-CoV-2) - an update on the status. *Infection, Genetics and Evolution, 83*, 1-10.

Acter, T., Uddin, N., Das, J., Akhter, A., Choudhury, T. R., & Kim, S. (2020). Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency. *Science of The Total Environment, 730*, 1-19.

BBC. (2020). *Coronavirus: Wuhan in first virus cluster since end of lockdown 2020*. Retrieved from https://www.bbc.com/news/world-asia-china-52613138 (accessed 30 June 2020).

Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour, 4*, 588-596.

Cao, J., Jiang, X., & Zhao, B. (2020). Epidemic Prediction of COVID-19. *Advances in Biotechnology & Microbiology, 15*(4), 71-83.

Charles, V., & Gherman, T. (2013). Achieving competitive advantage through big data. Strategic implications. *Middle-East Journal of Scientific Research, 16*(8), 1069-1074.

Charles, V., & Gherman, T. (2018). Big Data Analytics and Ethnography: Together for the Greater Good. In A. Emrouznejad & V. Charles (Eds.), *Big Data for the Greater Good* (pp. 19-33). Springer.

Charles, V., Tavana, M., & Gherman, T. (2015). The right to be forgotten-is privacy sold out in the big data age? *International Journal of Society Systems Science, 7*(4), 283-298.

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals, 135*, 1-6.

Dhama, K., Sharun, K., Tiwari, R., Dadar, M., Malik, Y. S., Singh, K. P., & Chaicumpa, W. (2020). COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. *Human Vaccines & Immunotherapeutics, 16*(6), 1232-1238.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases, 20*(5), 533-534.

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals, 134*, 1-5.

Graeden, E., Carlson, C., & Katz, R. (2020). Answering the right questions for policymakers on COVID-19. *The Lancet Global Health, 8*(6), e768-e769.

Hansun, S. (2017). New estimation rules for unknown parameters on holt-winters multiplicative method. *Journal of Mathematical and Fundamental Sciences, 49*(2), 127-135.

Hansun, S., Charles, V., Indrati, C. R., & Subanar. (2019). Revisiting the Holt-Winters' Additive Method for Better Forecasting. *International Journal of Enterprise Information Systems, 15*(2), 43-57.

Hasan, N. (2020). Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model. *Internet of Things, 11*, 1-10.

JHU CSSE. (2020). *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University 2020*. Retrieved from https://github.com/CSSEGISandData/COVID-19 (accessed 15 June 2020).

Kırbaş, İ., Sözen, A., Tuncer, A. D., & Kazancıoğlu, F. Ş. (2020). Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos, Solitons & Fractals, 138*, 1-7.

Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., & Hsueh, P.-R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents, 55*(3), 1-9.

Leung, K., Wu, J. T., Liu, D., & Leung, G. M. (2020). First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet, 395*, 1382-1393.

Li, H., Liu, S.-M., Yu, X.-H., Tang, S.-L., & Tang. C.-K. (2020). Coronavirus disease 2019 (COVID-19): current status and future perspectives. *International Journal of Antimicrobial Agents, 55*(5), 1-8.

Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., Smoot, J., Gregg, A. C., Daniels, A. D., Jervey, S., & Albaiu, D. (2020). Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Central Science, 6*(3), 315-331.

Mandal, M., Jana, S., Nandi, S. K., Khatua, A., Adak, S., & Kar, T. K. (2020). A model based study on the dynamics of COVID-19: Prediction and control. *Chaos, Solitons & Fractals, 136*, 1-12.

Mei, X., Lee, H.-C., Diao, K., Huang, M., Lin, B., Liu, C., et al. (2020). Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. *Nature Medicine*.

Park, M., Thwaites, R. S., & Openshaw, P. J. M. (2020). COVID-19: Lessons from SARS and MERS. *European Journal of Immunology, 50*(3), 308-311.

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS One, 15*(3), 1-8.

Pharmaceutical Technology. (2020). *COVID-19 concerns increase in May following easing of lockdowns 2020.* Retrieved from https://www.pharmaceutical-technology.com/news/covid-19-concerns-increase-in-may-2020/ (accessed 30 June 2020).

Qi, H., Xiao, S., Shi, R., Ward, M. P., Chen, Y., Tu, W., et al. (2020). COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. *Science of The Total Environment, 728*, 1-6.

Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C., & Roser, M. (2020). *Coronavirus (COVID-19) Cases*. Retrieved from https://ourworldindata.org/covid-cases (accessed 30 June, 2020).

Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling, 5*, 271-281.

Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). *Coronavirus Pandemic (COVID-19).* Retrieved from https://ourworldindata.org/coronavirus (accessed 30 June 2020).

Sabat, I., Neuman-Böhme, S., Varghese, N. E., Barros, P. P., Brouwer, W., van Exel, J., et al. (2020). United but divided: Policy responses and people's perceptions in the EU during the COVID-19 outbreak. *Health Policy*, 1-10.

Singhal, A., Singh, P., Lall, B., & Joshi, S. D. (2020). Modeling and prediction of COVID-19 pandemic using Gaussian mixture model. *Chaos, Solitons & Fractals, 138*, 1-8.

Spinelli, A., & Pellino, G. (2020). COVID-19 pandemic: perspectives on an unfolding crisis. *British Journal of Surgery, 107*(7), 785-787.

Swapnarekha, H., Behera, H. S., Nayak, J., & Naik, B. (2020). Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. *Chaos, Solitons & Fractals, 138*, 1-15.

The New York Times. (2020). Reopenings mark a new phase: global' trial-and-Error' played out in lives. *The New York Times*. Retrieved from https://www.nytimes.com/2020/05/07/world/europe/coronavirus-reopening-costs.html (accessed 17 May 2020).

'The race against COVID-19'. (2020). *Nature Nanotechnology, 15*(4), 239-240.

The Guardian. (2020). 'One big wave' – why the Covid-19 second wave may not exist. *The Guardian*. Retrieved from https://www.theguardian.com/world/2020/jul/29/one-big-wave-why-the-covid-19-second-wave-may-not-exist-coronavirus (accessed 29 July 2020).

Tiwari, S., Kumar, S., & Guleria, K. (2020). Outbreak Trends of Coronavirus Disease–2019 in India: A Prediction. *Disaster Medicine and Public Health Preparedness*, 1-6.

Torrealba-Rodriguez, O., Conde-Gutiérrez, R. A., & Hernández-Javier, A. L. (2020). Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. *Chaos, Solitons & Fractals, 138*, 1-8.

Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things, 11*, 1-16.

Wangping, J., Ke, H., Yang, S., Wenzhe, C., Shengshu, W., Shanshan, Y., Jianwei, W., Fuyin, K., Penggang, T., Jing, L., Liao, L., & Yao, H. (2020). Extended SIR Prediction of the Epidemics Trend of COVID-19 in Italy and Compared With Hunan, China. *Frontiers in Medicine, 7*, 1-7.

WHO. (2020). *WHO Coronavirus Disease (COVID-19) Dashboard 2020*. Retrieved from https://covid19.who.int/ (accessed 30 June 2020).

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., et al. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *The British Medical Journal*, 1-24.

Xie, M., & Chen, Q. (2020). Insight into 2019 novel coronavirus — An updated interim review and lessons from SARS-CoV and MERS-CoV. *International Journal of Infectious Diseases, 94*, 119-124.

Yan, L., Zhang, H.-T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., et al. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence, 2*, 283-288.

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., et al. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease, 12*(3), 165-174.