

Hull-WEMA: A Novel Zero-Lag Approach in the Moving Average Family, with an Application to COVID-19

Seng Hansun

*Informatics Department, Universitas Multimedia Nusantara
Tangerang, Indonesia*

Email: seng.hansun@lecturer.umn.ac.id

Vincent Charles*

*School of Management, University of Bradford
Bradford BD7 1DP, UK*

Email: C.Vincent3@bradford.ac.uk

Tatiana Gherman

*Faculty of Business and Law, University of Northampton
Northampton NN1 5PH, UK*

Email: Tatiana.Gherman@northampton.ac.uk

Vijayakumar Varadarajan

*School of Computer Science and Engineering, The University of New South Wales
Sydney, Australia*

Email: v.varadarajan@unsw.edu.au

* Corresponding Author

Hull-WEMA: A Novel Zero-Lag Approach in the Moving Average Family, with an Application to COVID-19

Abstract

The Moving Average (MA) is undeniably one of the most popular forecasting methods in time series analysis. In this study, we consider two variants of MA, namely the Weighted Exponential Moving Average (WEMA) and the Hull Moving Average (HMA). WEMA, which was introduced in 2013, has been widely used in different scenarios but still suffers from lags. To address this shortcoming, we propose a novel zero-lag Hull-WEMA method that combines HMA and WEMA. We apply and compare the proposed approach with HMA and WEMA by using COVID-19 time series data from ten different countries with the highest number of cases on the last observed date. Results show that the new approach achieves a better accuracy level than HMA and WEMA. Overall, the paper advocates a white-box forecasting method, which can be used to predict the number of confirmed COVID-19 cases in the short run more accurately.

Keywords

Time Series Forecasting; Moving Average; HMA; WEMA; Hull-WEMA; White-Box Model; COVID-19; Python 3.

1. Introduction

Time series analysis has become crucial in many different fields, especially for strategic decision-making. It involves time series data and has two main usages: first, to find the basic structure or pattern in the observed data, and second, to fit a model of the observed data for future prediction (Deb et al., 2017; Hansun, 2016). Furthermore, time series forecasting has also become an essential branch of big data analysis (Shen et al., 2020). There are many forecasting methods that have been developed in the literature, but generally, they can be classified into two big groups, *i.e.*, the traditional statistical forecasting methods and the advanced forecasting methods that incorporate different soft computing methods (Hansun and Subanar, 2016). Table 1 compares the characteristics of traditional and advanced forecasting methods.

Table 1

Traditional Versus Advanced Forecasting Methods

Traditional Forecasting Methods	Advanced Forecasting Methods
Employs statistical methods	Employs soft computing methods, such as machine learning
Simple to complex analysis	Complex analysis
Higher interpretability	Lower interpretability
Easy to be implemented	More difficult to be implemented

Small number of parameters	High number of parameters
Use many simplifying assumptions	Do not use many simplifying assumptions
Static	Dynamic and evolving
White box models	Black box models ⁺
Low computational cost	High computational cost ⁺⁺

⁺Interested readers may find a comparison table between white box and black box models in Hansun et al. (2021).

⁺⁺Interested readers may find comparison results between traditional statistical methods and machine learning models in the M3 competition (Makridakis et al., 2018).

In view of traditional forecasting methods, which become the main focus of this study, the Moving Average (MA) family of methods undeniably becomes one of the most popular methods used in different scenarios. The first usage of MA can be traced back to the early 1900s (Raudys and Pabarskaite, 2018). Also known by other aliases such as ‘autoregressive models’ (in the statistical field), ‘low-pass filters’ (in signal processing), ‘sliding window’, and ‘exponential smoothing’ (in the later development of MA methods) (Raudys and Pabarskaite, 2018), the method has been improved and developed to tackle various problems, which resulted in the birth of different variants of MA methods. The most basic and the earliest type is the Simple Moving Average (SMA) method, which simply takes the average of several historical data without applying any weighting factor for each data point (Svetunkov and Petropoulos, 2018). Since the most recent data will usually have greater effects on future values, the Weighted Moving Average (WMA) method was introduced, which applies the same basic formula of SMA, but with the application of linear weighting factor for each data point in the dataset (Hansun and Bonar Kristanda, 2017). In the next development phase, the Exponential Moving Average (EMA) method was introduced. It uses the same formulas as the WMA, but rather than using a linear function as the weighting factor calculation, it uses an exponential function (Bonar Kristanda and Hansun, 2019). EMA, which is best known as the exponential smoothing method, is widely accepted by the scientific community and applied in a wide range of fields, as can be seen in Anthony and Anggono (2019), Kim et al. (2020), and Rao (2020).

A novel approach to a ‘hybrid moving average’ method, which combines the weighting factor calculation in WMA with the general procedure of EMA, was introduced in 2013 by Hansun (2013). The introduced method is known as the Weighted Exponential Moving Average (WEMA) method and has been proven to excel the WMA and EMA method in terms of accuracy level. Moreover, since the introduction of this approach, it has attracted the attention of many researchers to apply the method in their respective fields. The robustness of the WEMA method has been tested in different kinds of scenarios and types of implementation, such as in the prediction of major Forex currency pairs (Hansun and Bonar Kristanda, 2018), prediction of Indonesia’s export and import values and volumes (Hansun et al., 2018), and ASEAN capital markets forecasting (Hansun et al., 2019). Moreover, Chen et al. (2015) also proposed a new differential secret key generation scheme based on the WEMA method, called the WEMA differential secret key generation (WEMA-DSKG). They found that the proposed method had a higher secret bit rate and entropy, with a lower bit mismatch rate. In the medical field, Rashidi Khazaei et al. (2019) proposed a hybrid prediction model based on multi-layer perceptron, linear regression, and the WEMA method that can predict the function of the transplanted kidney in a long-term care process. They found that the hybrid model could give reliable prediction values for estimated glomerular filtration rate (eGFR) in the routine daily care of kidney transplant patients.

Despite its acceptance and advantages, the WEMA method still suffers from several weaknesses. Like other MA methods, it experiences ‘lag’ where it shows a previous trend instead of the new ‘desired’ one (Raudys et al., 2013). Therefore, some researchers focused their study on reducing the lag in the MA methods, efforts which gave birth to the zero-lag approaches. One of the most famous examples of the zero-lag approaches is the Hull Moving Average (HMA) introduced by Hull (Raudys and Pabarskaite, 2018). It has been applied in different scenarios, such as for stock price series smoothing (Raudys, 2014) and trading systems (Di Lorenzo, 2012). Usaratniwart et al. (2017) developed a new smoothing technique based on HMA. The technique is called ‘adaptive enhanced linear exponential smoothing’ (AELES) and has the ability to mitigate the voltage fluctuation. Similarly, Ali et al. (2019) applied HMA to extenuate voltage fluctuation in electric vehicles (EV). Together with the gravitational search algorithm (GSA), they found that the proposed method effectively smoothed EV batteries' voltage fluctuations.

In this study, we are concerned with taking the advantages of both WEMA and HMA methods to build a new zero-lag approach in the MA family, called the Hull-WEMA method. Moreover, the proposed method will be compared with several other popular forecasting methods in the MA family, such as WEMA and HMA. The remainder of the paper is structured as follows. In the next section 2, we provide a brief discussion of the proposed method's two building block methods, *i.e.*, the HMA and WEMA. Then, section 3 will focus on the introduction of the proposed forecasting method, the Hull-WEMA method. The experimental and comparison results of the proposed method with other MAs methods will be given in section 4. Finally, section 5 comprises the concluding remarks.

2. The Building Block Methods

In this section, a description of two methods, namely WEMA and HMA, that constitute the building blocks of our proposed method, is given. Both methods have been widely accepted and used in the literature. We will also give the basic codes for both methods, which are written in Python programming language.

2.1 Weighted Exponential Moving Average

Hansun (2013) introduced a new hybrid approach to the MA forecasting method for time series analysis. Since it combines two conventional MA, namely WMA and EMA, the method was called the ‘Weighted Exponential Moving Average’ or WEMA. The method takes the advantages of both methods where some recent data will be considered in the prediction of future values using the EMA approach. There are three steps in the WEMA procedure, which are detailed below:

- 1) Calculate the WMA with period n using Eq. (1):

$$WMA_t = \frac{\sum_{t=k-n+1}^k w_t A_t}{\sum_{t=k-n+1}^k w_t}, \quad (1)$$

where n is the period or span number, k is the relative position of the current data point being considered, A_t is the actual value at time t , and w_t is the linear weight at time t .

- 2) Calculate the WEMA by using the EMA formula as shown in Eq. (2):

$$WEMA_t = \alpha \cdot WMA_t + (1 - \alpha) \cdot WEMA_{t-1}, \quad (2)$$

where α is the smoothing constant taking a value between 0 and 1, WMA_t is the base value obtained from Step 1, and $WEMA_t$ is the predicted value at time t . For $t = 1$, $WEMA_t = A_t$.

- 3) Back to Step 1 until all data points have been visited.

Although there is a slightly different formula for the WEMA method in the literature, the above procedure will be used in this study. Figure 1 shows the implemented codes for WEMA in Python language.

```
# Weighted Exponential Moving Average
def WEMA(src, period, alpha):
    wema = [None] * len(src)
    wmaFull = WMA(src, period)
    wema[period-2] = src[period-2]
    for i in range(len(src)-period+1):
        wema[i+period-1] = ((alpha * wmaFull[i+period-1]) + ((1 - alpha) * wema[i+period-2]))
    return wema
```

Fig. 1. WEMA implementation in Python.

2.2 Hull Moving Average

To solve the old-time dilemma of lag in MA, Hull (Zakamulin, 2017) introduced the HMA method. This method uses a combination of three WMAs with different sizes of period. It is considered as one of the best approaches to reducing the average lag time of MA, while simultaneously improving the smoothing (Hull, n.d.). The procedure of HMA can be described in three steps, as follows:

- 1) Calculate the WMA with period $n/2$ and multiply it by 2.
- 2) Calculate the WMA with period n and subtract it from Step 1.
- 3) Calculate the WMA with period \sqrt{n} using the results from Step 2.

The above procedure can be written as a mathematical formula, as shown in Eq. (3) (Kolkova, 2018; Oyewola et al., 2021):

$$HMA = WMA_{\sqrt{n}} \left(2 \times WMA_{\frac{n}{2}}(A) - WMA_n(A) \right), \quad (3)$$

where WMA_n , $WMA_{\frac{n}{2}}$, and $WMA_{\sqrt{n}}$ are the results for WMA with periods of 'integer' n , $\frac{n}{2}$, and \sqrt{n} respectively. Furthermore, A is the actual observed data, and HMA represents the predicted results. Figure 2 shows the implemented codes for HMA in Python language.

```
# Hull Moving Average
def HMA(src, period):
    hma = [None] * len(src)
    inp = [None] * len(src)
    wmaHalf = WMA(src, math.floor(period/2))
    wmaFull = WMA(src, period)
    inp[period-2] = src[period-2]
    for i in range(len(src)-period+1):
        inp[i+period-1] = (2 * wmaHalf[i+period-1]) - wmaFull[i+period-1]
    hma[(period-2):] = WMA(inp[(period-2):], math.floor(math.sqrt(period)))
    return hma
```

Fig. 2. HMA implementation in Python.

3. Hull-WEMA and Error Criteria

In this section, we introduce and describe our proposed approach to the zero-lag method in the MA family. Moreover, we also briefly describe the forecast error criteria being implemented in this study, namely the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Scaled Error (MASE).

3.1 Hull Weighted Exponential Moving Average

As suggested by its name, Hull-WEMA, our proposed novel approach for the zero-lag MA combines the HMA and the WEMA methods. The motivation behind this approach is quite simple. Since the HMA method could reduce the average lag time in the given time series data, we first use the HMA method described in Section 2.2 to get the base values. Next, the base values obtained will be used as the input in the second step of the WEMA method, as explained in Section 2.1. By combining both methods, we could get the forecasting results that follow the procedure of WEMA, while reducing the average lags at the same time. A more detailed, yet simple procedure of the proposed method, the Hull-WEMA, is as follows:

- 1) Calculate the HMA using Eq. (3) for the given time series data and period.
- 2) Calculate the predicted values using Eq. (4):

$$Hull\ WEMA_t = \alpha \cdot HMA_t + (1 - \alpha) \cdot Hull\ WEMA_{t-1}, \quad (4)$$

where α is the smoothing constant that takes a value between 0 and 1, HMA_t is the base value obtained from Step 1, and $Hull\ WEMA_t$ is the predicted value at time t . For $t = 1$, $Hull\ WEMA_t = A_t$.

- 3) Back to Step 1 until all data points have been visited.

The implemented codes for the Hull-WEMA method in Python language are shown in Figure 3. Moreover, the complete codes can be retrieved from GitHub at <https://github.com/senghansun/Hull-WEMA> (Hansun, 2020).

```
# Hull Weighted Exponential Moving Average
def HullWEMA(src, period, alpha):
    hullWema = [None] * len(src)
    hma = HMA(src, period)
    hullWema[period-2] = src[period-2]
    for i in range(len(src)-period+1):
        hullWema[i+period-1] = ((alpha * hma[i+period-1]) + ((1 - alpha) * hullWema[i+period-2]))
    return hullWema
```

Fig. 3. Hull-WEMA implementation in Python.

3.2 Error Criteria

Two popular forecast error criteria used in this study are MAPE and MASE. Both criteria are included as unit-free [or scale-independent (Kim and Kim, 2016)] forecast error measurements, which means that they can serve as effective measures of model accuracy, typically against datasets with vastly different sizes (Giri et al., 2020; Kuo, 2019).

MAPE is one of the most widely used error evaluation methods in the literature. It shows the percentage value of absolute error compared to the actual ones (Hansun et al., 2019). MAPE can be represented as in Eq. (5) (Prapcoyo, 2018; Saoud and Al-Marzouqi, 2020):

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{A_t - F_t}{A_t} \right|. \quad (5)$$

MASE is another forecast error measurement proposed by Hyndman and Koehler in 2006 (Hyndman and Koehler, 2006). It uses a naïve forecast approach and the in-sample Mean Absolute Error (MAE) to scale the absolute error value (Hansun et al., 2019). For non-seasonal time series data, it can be represented as in Eq. (6) (Hyndman and Koehler, 2006; Jiao et al., 2020):

$$MASE = \frac{1}{N} \sum_{t=1}^N \left| \frac{A_t - F_t}{\frac{1}{N-1} \sum_{t=2}^N |A_t - A_{t-1}|} \right|, \quad (6)$$

where N is the total amount of data in the given time series, A_t is the actual value at time t , and F_t is the forecasted value at time t .

4. Experimental Results

To study the proposed method's accuracy level, we performed experiments using the Hull-WEMA method and several other MA methods on COVID-19 data. More precisely, we tried to predict the values of confirmed cases of COVID-19 disease in several countries. The reason behind choosing to work with COVID-19 data lies in the contemporaneity of the disease. COVID-19 is a significant public health issue, which is caused by a virus called the *Severe Acute Respiratory Syndrome Coronavirus-2* (Lai et al., 2020). Moreover, due to its massive global spreading and range of impacts (from individual to organisational and national level), the World Health Organization (WHO) declared the disease as a pandemic in March 2020 (Spinelli and Pellino, 2020). Its future course is still unknown and research efforts are constantly being made to not only identify a vaccine, but also to predict the number of future infections and deaths, as well as fight the negative consequences affecting economies worldwide. We therefore position this paper as a contribution towards developing a better forecasting method, which can be used to predict the number of confirmed COVID-19 cases more accurately, in the short run.

Section 4.1 gives a brief description of the data being used in this study. The prediction results using several MA methods are given in Section 4.2, and the accuracy level for each method, which is defined by its error calculation, is given in Section 4.3.

4.1 Data Collection

The primary data being used in this study are the time series data of confirmed cases of COVID-19 disease. We collected the data from a GitHub repository managed by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) (JHU CSSE, 2020). The repository contains the confirmed, recovered, and death cases of COVID-19, which are used by an online real-time interactive dashboard for COVID-19 developed by the JHU team (Dong et al., 2020). We use the time series data for ten countries that had the highest number of confirmed cases on the last observed date, *i.e.*, 8 October 2020; the data is provided in Table 2. Figure 4 shows the Python codes to get the top ten countries, while Figure 5 plots the COVID-19 confirmed cases for those countries.

Table 2

COVID-19 Time Series Data Being Used in the Study

No	Country	Date Range	Number of Data Points (Days)
1	Argentina	3 Mar 2020 – 8 Oct 2020	220
2	Brazil	26 Feb 2020 – 8 Oct 2020	226
3	Colombia	6 Mar 2020 – 8 Oct 2020	217
4	India	30 Jan 2020 – 8 Oct 2020	253
5	Mexico	28 Feb 2020 – 8 Oct 2020	224
6	Peru	6 Mar 2020 – 8 Oct 2020	217
7	Russia	31 Jan 2020 – 8 Oct 2020	252
8	South Africa	5 Mar 2020 – 8 Oct 2020	218
9	Spain	1 Feb 2020 – 8 Oct 2020	251
10	US	22 Jan 2020 – 8 Oct 2020	261
Total			2,339

```
## last-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
fig = plt.figure(figsize=(20,10))
ax = fig.add_subplot(111)
for i in range(len(countries)):
    if data["Selected_Global"][i]:
        y = data.iloc[i,3:-2]
        if np.isnan(data.index[i]):
            lbl = str(data.iloc[i,0])
        else:
            lbl = str(data.iloc[i,0]) + " - " + str(data.index[i])
        ax.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)

# title, label, and legend
ax.set_title("COVID-19 Confirmed - Global", fontsize=18, fontweight='bold')
ax.set_xlabel('Time', fontsize=15)
ax.set_ylabel('Number of confirmed COVID-19 cases', fontsize=15)
ax.legend()

fig.savefig('Top Ten Countries.jpg', bbox_inches = 'tight')
```

Fig. 4. Python implementation to get the top ten countries.

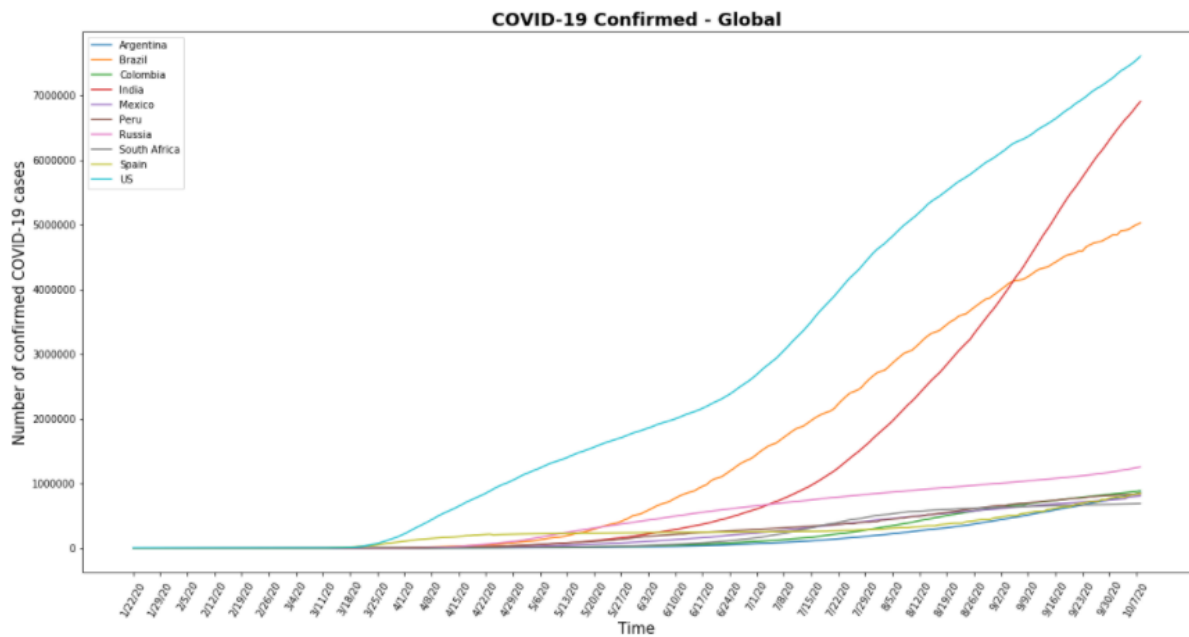


Fig. 5. The number of COVID-19 confirmed cases for the top ten countries considered in the study.

4.2 Prediction Results and Analysis

The prediction results for the top ten countries being considered in the study are given in this section. We show the prediction results obtained by using WEMA, HMA, and the proposed zero-lag approach, *i.e.*, the Hull-WEMA method. First, we split the datasets, as explained in Section 4.1, into train and test data with a ratio of 80:20. Table 3 shows the number of train and test data for each country with their respective periods.

Table 3

Train vs. Test Data Split

Country	Total Data Points (Days)	Train		Test	
		Date Range	Days	Date Range	Days
Argentina	220	3 Mar 20 – 25 Aug 20	176	26 Aug 20 – 8 Oct 20	44
Brazil	226	26 Feb 20 – 23 Aug 20	180	24 Aug 20 – 8 Oct 20	46
Colombia	217	6 Mar 20 – 25 Aug 20	173	26 Aug 20 – 8 Oct 20	44
India	253	30 Jan 20 – 18 Aug 20	202	19 Aug 20 – 8 Oct 20	51
Mexico	224	28 Feb 20 – 24 Aug 20	179	25 Aug 20 – 8 Oct 20	45
Peru	217	6 Mar 20 – 25 Aug 20	173	26 Aug 20 – 8 Oct 20	44
Russia	252	31 Jan 20 – 18 Aug 20	201	19 Aug 20 – 8 Oct 20	51

South Africa	218	5 Mar 20 – 25 Aug 20	174	26 Aug 20 – 8 Oct 20	44
Spain	251	1 Feb 20 – 18 Aug 20	200	19 Aug 20 – 8 Oct 20	51
US	261	22 Jan 20 – 16 Aug 20	208	17 Aug 20 – 8 Oct 20	53

The purpose of the training phase is to find the best smoothing constant value (α), which is used by both the WEMA and Hull-WEMA methods. The best value is found by rerunning both methods several times (the default value is 100) that could give the lowest Mean Absolute Percentage Error (MAPE) as the error measurement criteria. Moreover, we also use seven data periods as the default input value for all applied methods in this study. Table 4 shows the best α value for each country found during the training phase.

Table 4

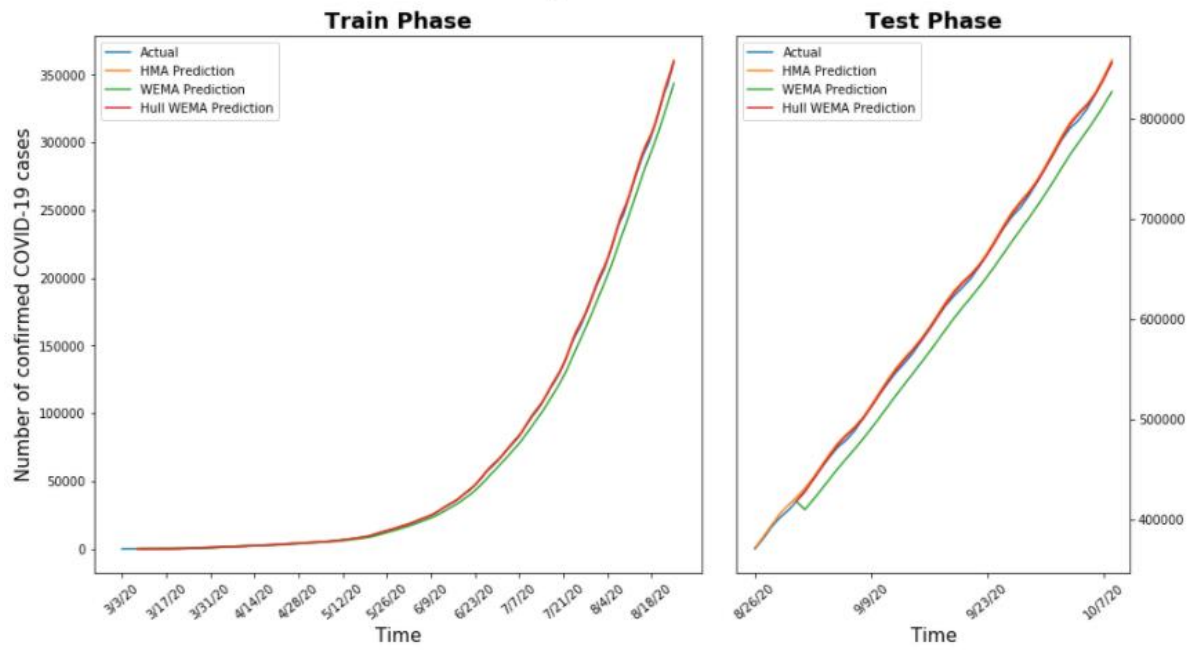
Best α for WEMA and Hull-WEMA

Country	WEMA		Proposed Hull-WEMA	
	Best α	MAPE	Best α	MAPE
Argentina	0.99	10.582578	0.85	1.367974
Brazil	0.99	12.432575	0.89	2.000314
Colombia	0.99	11.199843	0.84	1.329066
India	0.99	11.180390	0.83	1.655934
Mexico	0.99	10.713949	0.87	1.326896
Peru	0.99	10.359019	0.87	1.647792
Russia	0.99	10.117957	0.92	1.432328
South Africa	0.99	11.052603	0.87	1.566090
Spain	0.99	8.519508	0.99	1.885227
US	0.99	10.226317	0.99	1.917176

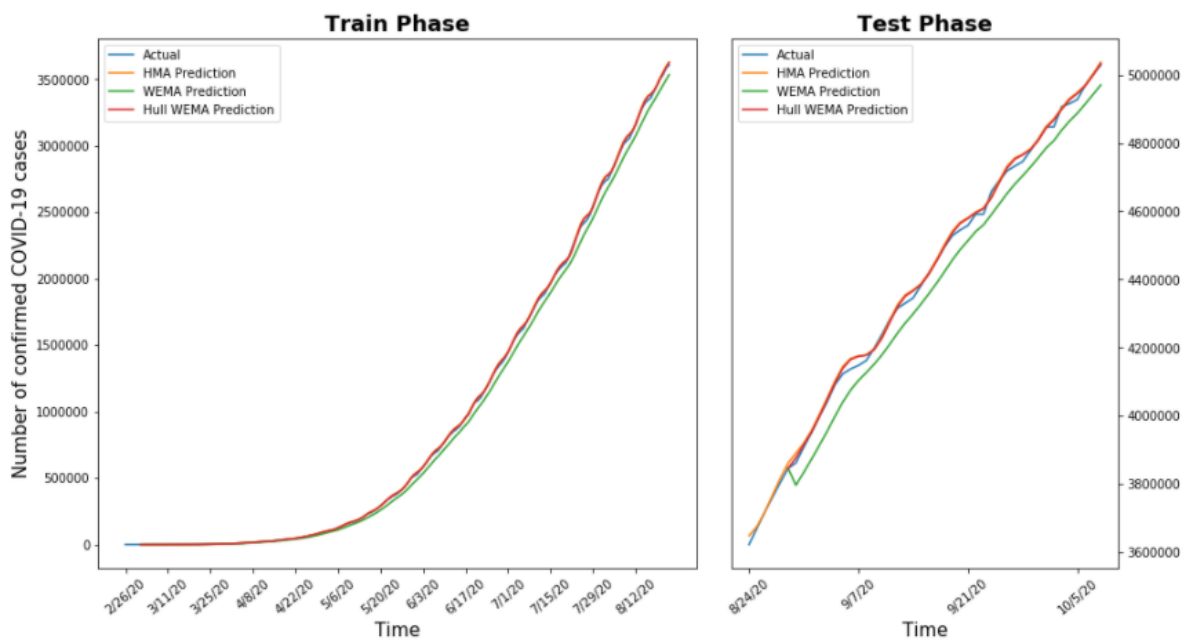
In the testing phase, we use the best α found in the training phase to run the algorithm for each respective method. Since there is no smoothing constant parameter in HMA, we use all the data points for the prediction without splitting them into train and test data. Meanwhile, in both WEMA and Hull-WEMA methods, the smoothing constant parameter will be used in the testing phase for future prediction of the time series data. To compare all methods' prediction results, we use two unit-free error measurements, namely MAPE and MASE, as explained in Section 3.2.

The prediction results for each country being considered in this study are given in Figure 6. The 'blue' line denotes the actual data, the 'orange' line represents the prediction results using HMA, the 'green' line shows the prediction results using WEMA, and the 'red' line is for the prediction results using Hull-WEMA. Moreover, the MAPE and MASE values for each method applied in the study are given in Table 5.

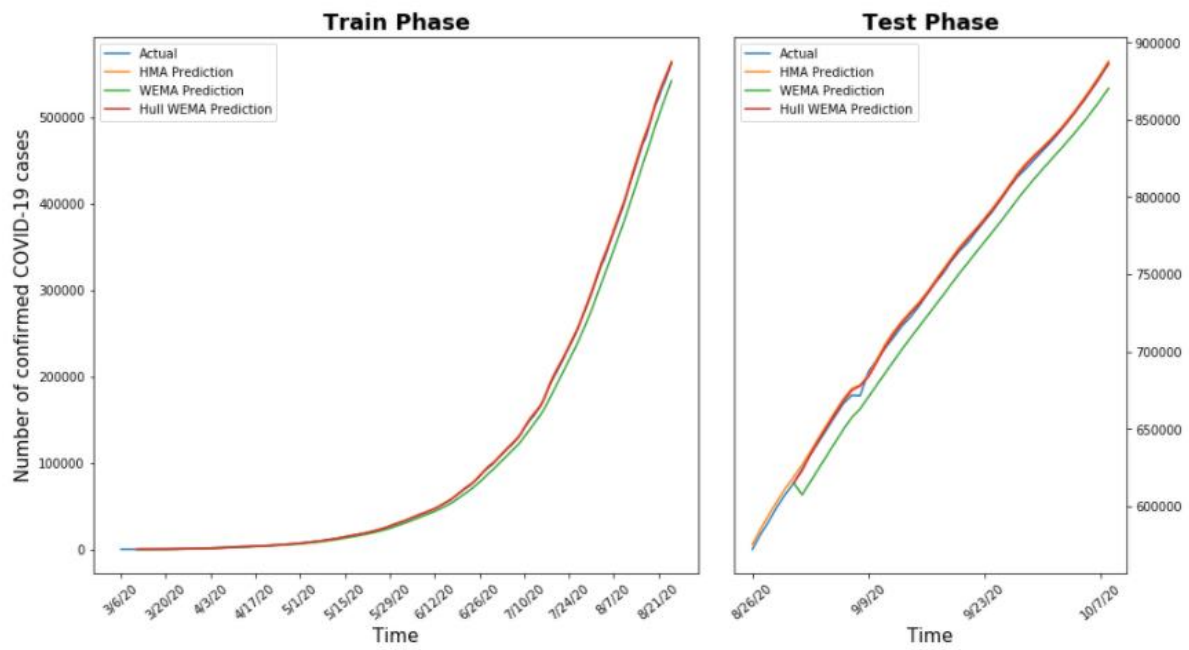
Argentina



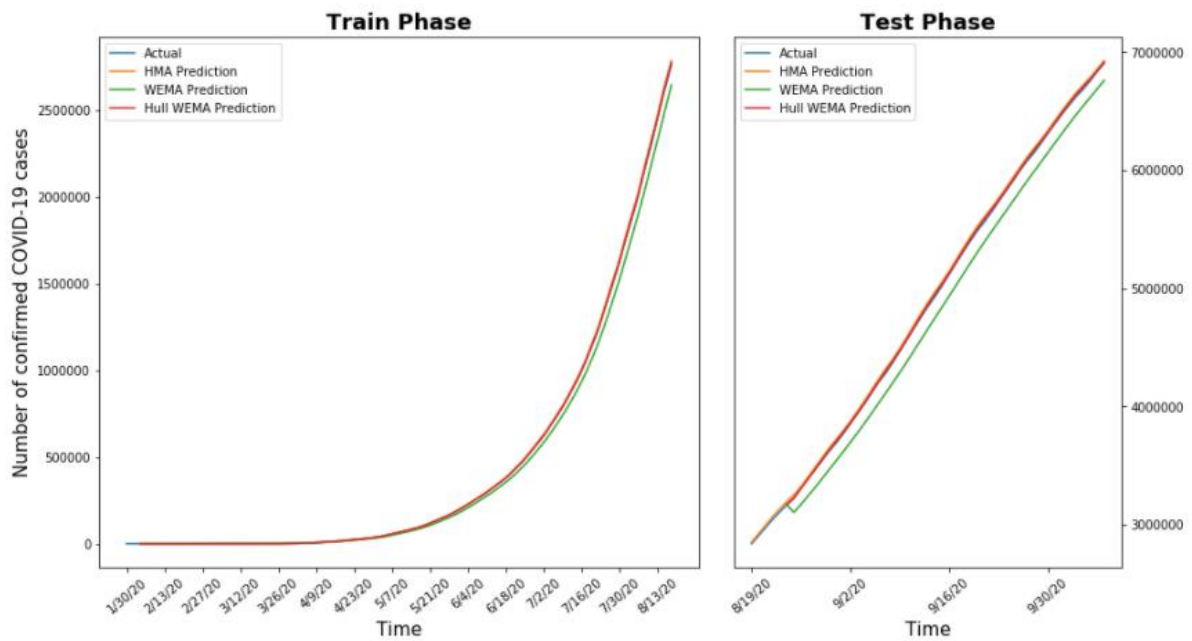
Brazil

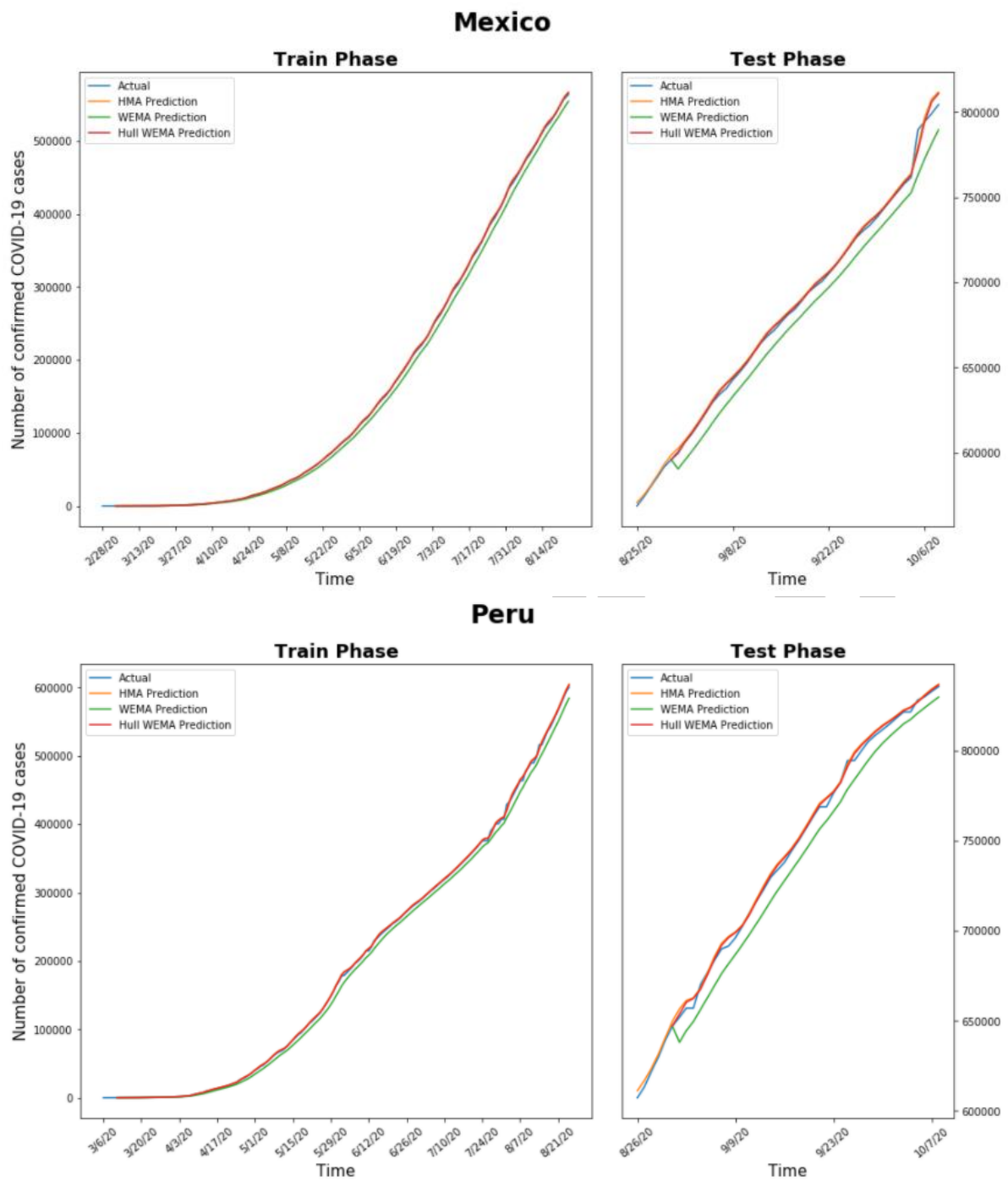


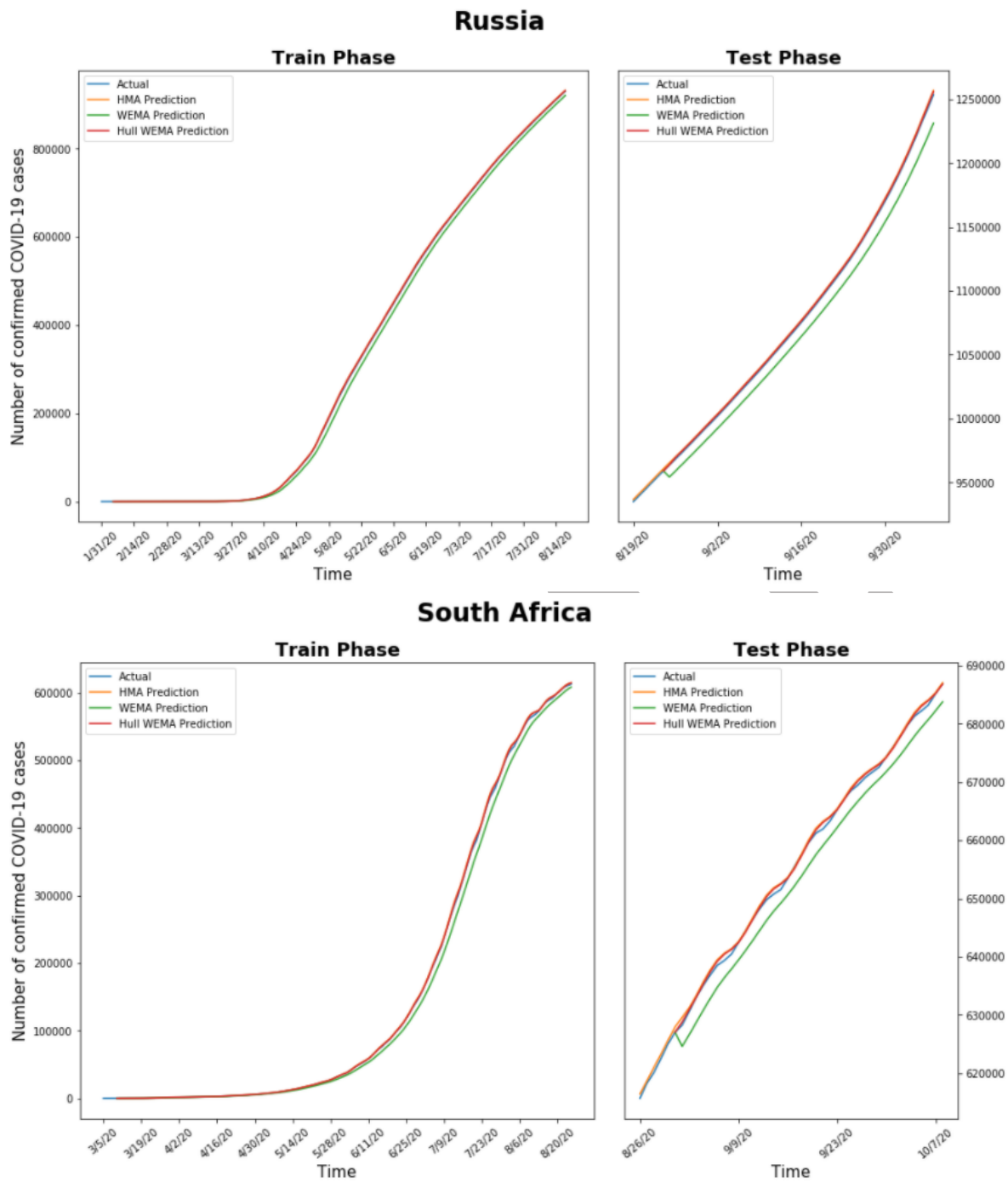
Colombia



India







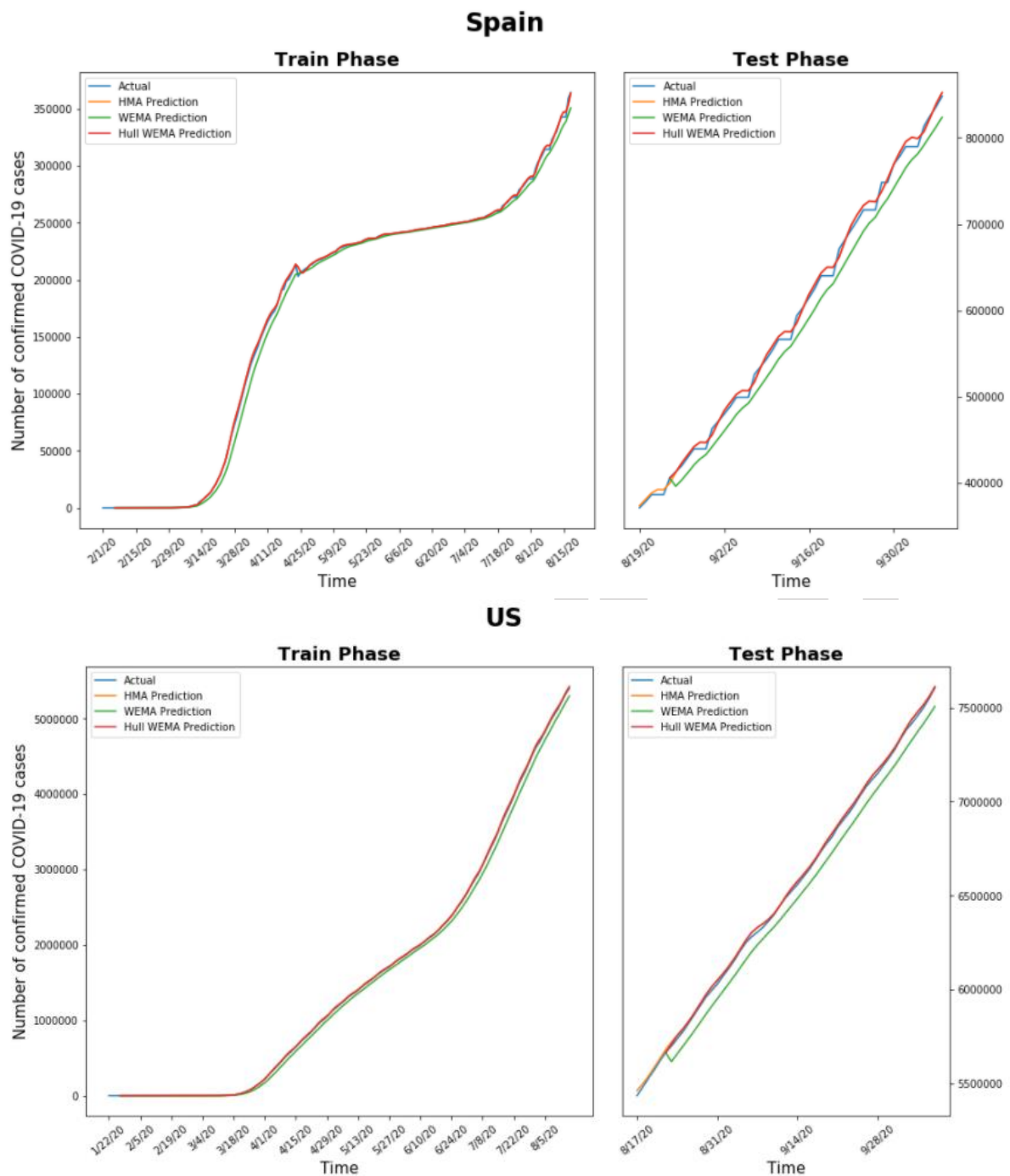


Fig. 6. Prediction results of COVID-19 confirmed cases using WEMA, HMA, and Hull-WEMA.

Table 5

MAPE and MASE Values in the Testing Phase

Country	WEMA		HMA		Proposed Hull-WEMA	
	MAPE	MASE	MAPE	MASE	MAPE	MASE
Argentina	3.692057	1.920786	1.444371	0.290716	0.281054	0.147563
Brazil	1.361809	1.986339	1.753350	0.344589	0.234135	0.343359
Colombia	1.920936	1.949919	1.362404	0.315808	0.177549	0.175244
India	3.423444	1.958746	1.738152	0.307172	0.200374	0.121326
Mexico	1.527995	1.942859	1.289052	0.336226	0.199625	0.264934
Peru	1.395531	2.027225	1.497031	0.388704	0.235260	0.340934
Russia	1.161244	1.911082	1.258953	0.311232	0.113969	0.185956
South Africa	0.487350	1.979444	1.366751	0.329459	0.061510	0.249731
Spain	3.251890	1.953809	1.679493	0.500925	0.913386	0.549944
US	1.252714	1.959763	1.561818	0.318247	0.196906	0.307004
AVERAGE	1.947497	1.958997	1.495138	0.344308	0.261377	0.268600

As shown in Table 5, the proposed approach, *i.e.*, the Hull-WEMA method, has the lowest average values for both MAPE and MASE. Intuitively, we could say that the proposed approach has a better accuracy level when compared to the WEMA and HMA methods due to its smaller error values. Figure 7 shows the scatter plots that compare the forecast error results for each method. It can be seen clearly from the plot that the proposed method gives lower MAPE and MASE values for each country considered in the study.

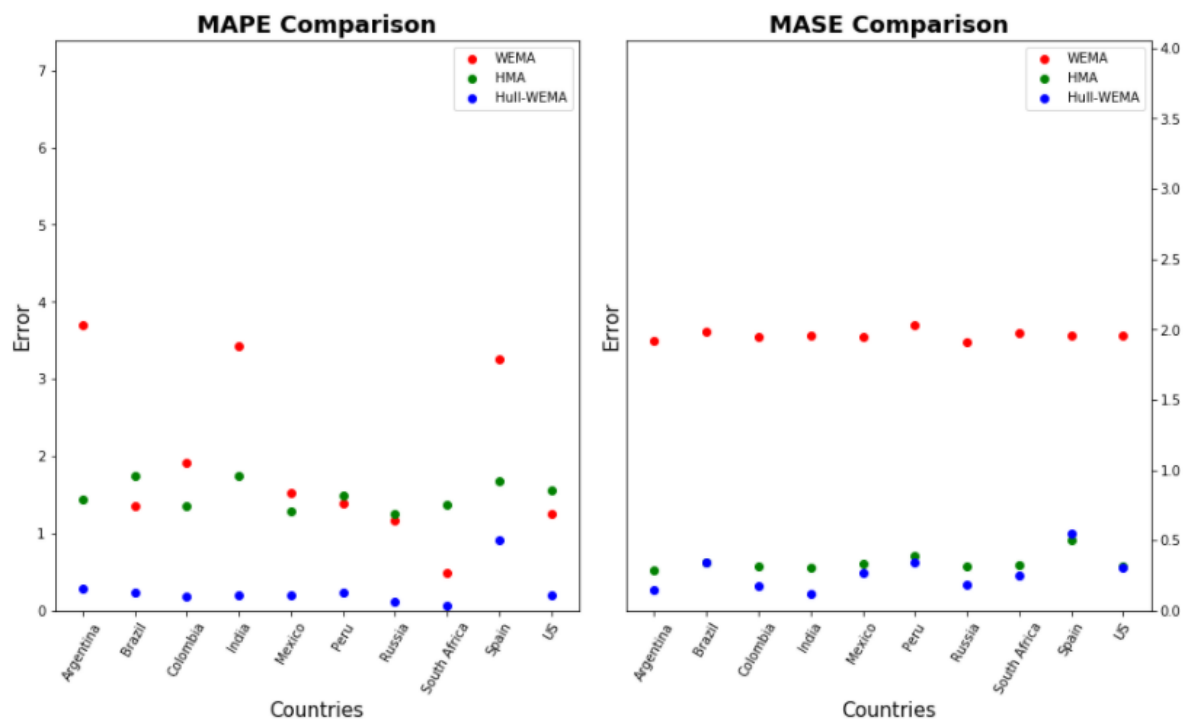


Fig. 7. Scatter plot of MAPE and MASE for each method.

Moreover, to confirm this result, we conducted a paired sample t -test using an add-in in Microsoft Excel called 'Data Analysis ToolPak'. First, we compare the MAPE and MASE results for WEMA versus Hull-WEMA, then similarly we compare the MAPE and MASE results for HMA versus Hull-WEMA. Table 6 shows the paired sample t -test results for all comparisons conducted. Since the two-tailed p -values for all comparisons are less than the significance level value at 0.05, we can conclude with 95% confidence that the difference in the means of MAPE and MASE values for WEMA (and HMA) versus Hull-WEMA are significantly different. Specifically, the MAPE and MASE values of the Hull-WEMA method are lower than those of WEMA and HMA.

Table 6

Paired Sample t-test Results

Comparison		Pearson Correlation	t -Stat	df	t Critical (1-tailed)	Sig. (1-tailed)	t Critical (2-tailed)	Sig. (2-tailed)
WEMA vs. Hull-WEMA	MAPE	0.55325	5.37	9	1.83311	0.00023	2.26216	0.00045
	MASE	0.38859	45.49	9	1.83311	2.998E-12	2.26216	5.995E-12
HMA vs. Hull-WEMA	MAPE	0.45522	17.41	9	1.83311	1.534E-08	2.26216	3.068E-08
	MASE	0.92007	3.23	9	1.83311	0.00515	2.26216	0.01031

5. Conclusions

In this study, we have described two popular MA methods, namely the WEMA and HMA methods. We took the advantages of both methods and proposed a new approach for the zero-lag method in the MA family, *i.e.*, the Hull-WEMA method. Then, we applied and compared the proposed approach with the WEMA and HMA methods using time series data of COVID-19 confirmed cases in ten different countries with the highest case number on the last observed date (*i.e.*, 8 October 2020). We found that the proposed approach could give better prediction results when compared to the WEMA and HMA methods.

As noted by Hansun et al. (2021), large amounts of data within the framework of big data are constantly being generated at an exponential rate, holding the potential to expand the opportunities for making better decisions at strategic, tactical, and operational levels (Charles, Tavana, & Gherman, 2015). The challenge, however, is to find ways to translate these data into meaningful knowledge (Charles & Gherman, 2013) in a timely manner, especially in the context of the COVID-19 pandemic. Recently, many studies have emerged that use soft computing methods, also known as black-box models, for COVID-19 prediction purposes. Nonetheless, as stated in Hansun et al. (2021), more sophisticated models are not necessarily better than white-box models. For example, black-box models, despite providing higher accuracy, have high computational complexity, lower interpretability or explainability, and lower transparency and accountability, among others. White-box models, on the other hand, have relatively lower accuracy levels, but are highly transparent, accountable, and explainable, with clarity around inner workings. In this sense, then, in the context in which the COVID-19 pandemic is not yet fully understood, white-box models become more reliable, as they provide

stakeholders with pieces of information that clearly show how the models behave and which variables precisely are influencing the predictions.

Overall, this paper advocates a white-box forecasting method, which can be used to predict the number of confirmed COVID-19 cases more accurately, in the short run. In this sense, we join the recent research efforts made by the community of researchers to assist governments, policymakers, and other relevant stakeholders alike by providing forecasts that can be used as a tool towards making better decisions and taking appropriate actions to contain or curb the spread of the coronavirus.

An avenue for future research is to test the acceptance and robustness level of the proposed approach in different fields and on different datasets. It is also proposed to conduct a performance analysis of the proposed approach with other zero-lag MA methods, such as the Zero Lag Moving Average (ZLMA), the Exponential Hull Moving Average (EHMA), and the MESA Adaptive Moving Average (MAMA) (Ehlers, 2002; Ehlers, n.d.; Raudys et al., 2013). Moreover, future research could employ qualitative approaches (Charles & Gherman, 2018) with a view to integrating the various stakeholders (Charles et al., 2019), to complement the analyses performed in this paper, so as to build a more solid picture of the COVID-19 pandemic and its behaviour.

Data Availability

All data being used in this study, including the Python codes for all applied methods, are available to be accessed by the public from the GitHub repository at <https://github.com/senghansun/Hull-WEMA>.

References

- Ali, A., Raisz, D., & Mahmoud, K. (2019). Voltage fluctuation smoothing in distribution systems with RES considering degradation and charging plan of EV batteries. *Electric Power Systems Research*, 176, 105933.
- Anthony, & Anggono, A. H. (2019). Investment Strategy Based on Exponential Moving Average and Count Back Line. *Review of Integrative Business and Economics Research*, 8(4), 153-161.
- Bonar Kristanda, M., & Hansun, S. (2019). MOODLE LMS Resources Prediction: Exponential Moving Average Approach. *International Journal of Advanced Trends in Computer Science and Engineering*, 8(4), 1308-1311.
- 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., Gherman, T., & Paliza, J. C. (2019). Stakeholder Involvement for Public Sector Productivity Enhancement: Strategic Considerations. *ICPE Public Enterprise Half-Yearly Journal*, 24(1), 77-86.
- 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.
- Chen, M., Jiang, T., & Zou, W. (2015). Differential physical layer secret key generation based on weighted exponential moving average. *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Cairns, QLD, Australia, pp. 1-4.

- Deb, C., Zhang, F., Yang, J., Lee, S. E., Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74, 902-924.
- Di Lorenzo, R. (2012). The Hull Algorithm. In: *How to Make Money by Fast Trading. Perspectives in Business Culture* (pp. 87-93). Milano: Springer.
- 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.
- Ehlers, J. F. (2002). Zero-Lag Data Smoothers. *Stocks & Commodities*, 20(7), 26-31.
- Ehlers, J. (n.d.). *MAMA - The Mother of Adaptive Moving Averages*. Retrieved from <https://www.mesasoftware.com/papers/MAMA.pdf> (accessed 15 October 2020).
- Giri, S., Purkayastha, S., Hazra, S., Chanda, A., Das, I., & Das, S. (2020). Prediction of monthly Hilsa (*Tenualosa ilisha*) catch in the Northern Bay of Bengal using Bayesian structural time series model. *Regional Studies in Marine Science*, 39, 101456.
- Hansun, S. (2013). A new approach of moving average method in time series analysis. *2013 Conference on New Media Studies*, Tangerang, Indonesia, pp. 1-4.
- Hansun, S. (2016). A New Approach of Brown's Double Exponential Smoothing Method in Time Series Analysis. *Balkan Journal of Electrical & Computer Engineering*, 4, 75-78.
- Hansun, S. (2020). *Hull-WEMA: A New Zero-Lag Approach in MAs family 2020*. Retrieved from <https://github.com/senghansun/Hull-WEMA> (accessed 12 October 2020).
- Hansun, S., & Bonar Kristanda, M. (2017). Performance analysis of conventional moving average methods in forex forecasting. *2017 International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, Yogyakarta, Indonesia, pp. 11-17.
- Hansun, S., & Bonar Kristanda, M. (2018). Major FX Forecasting with Hybrid Moving Average Approach. *International Journal of Engineering & Technology*, 7(4.40), 108-111.
- Hansun, S., Bonar Kristanda, M., & Winarno, P. M. (2018). Indonesia's Export-Import Prediction: A Hybrid Moving Average Approach. *Journal of Engineering and Science Research*, 2(6), 37-43.
- Hansun, S., Bonar Kristanda, M., & Winarno, P. M. (2019). Big 5 ASEAN capital markets forecasting using WEMA method. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 17(1), 314-319.
- Hansun, S., Charles, V., Gherman, T., Subanar, & Rini-Indrati, C. (2021). A Tuned Holt-Winters White-Box Model for COVID-19 Prediction. *International Journal of Management and Decision Making*. In press.
- 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.
- Hansun, S., & Subanar, S. (2016). H-WEMA: A New Approach of Double Exponential Smoothing Method. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 14(2), 772-777.
- Hull, A. (n.d.). *How to reduce lag in a moving average*. Retrieved from <https://alanhull.com/hull-moving-average> (accessed 12 September 2020).
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.
- Jiao, X., Li, G., & Chen, J. L. (2020). Forecasting international tourism demand: a local spatiotemporal model. *Annals of Tourism Research*, 83, 102937.
- 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 10 October 2020).

- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669-679.
- Kim, T., Solanki, V. S., Baraiya, H. J., Mitra, A., Shah, H., & Roy, S. (2020). A Smart, Sensible Agriculture System Using the Exponential Moving Average Model. *Symmetry*, 12(3), 457.
- Kolkova, A. (2018). Indicators of Technical Analysis on the Basis of Moving Averages as Prognostic Methods in the Food Industry. *Journal of Competitiveness*, 10(4), 102-119.
- Kuo, K. (2019). DeepTriangle: A Deep Learning Approach to Loss Reserving. *Risks*, 7, 97.
- 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), 105924.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3), e0194889.
- Oyewola, D. O., Augustine, A. F., Dada, E. G., & Ibrahim, A. (2021). Predicting Impact of COVID-19 on Crude Oil Price Image with Directed Acyclic Graph Deep Convolutional Neural Network. *Journal of Robotics and Control*, 2(2), 103-109.
- Prapcoyo, H. (2018). PERAMALAN JUMLAH MAHASISWA MENGGUNAKAN MOVING AVERAGE. *Telematika: Jurnal Informatika dan Teknologi Informasi*, 15(1), 67-76.
- Rao, N. M. (2020). Stock Option Risk Management using Exponential Moving Average Method – A New Approach. *International Conference on Technology and Business Management* (Online), pp. 1–15.
- Rashidi Khazaei, P., Jamshid Bagherzadeh, M., Niazkhani, Z., & Pirnejad, H. (2019). Predicting the function of transplanted kidney in long-term care processes: Application of a hybrid model. *Journal of Biomedical Informatics*, 91, 103116.
- Raudys, A. (2014). Optimal negative weight moving average for stock price series smoothing. *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER)*, London, pp. 239-246.
- Raudys, A., Lenčiauskas, V., & Malčius, E. (2013). Moving Averages for Financial Data Smoothing. In T. Skersys, R. Butleris, & R. Butkiene (Eds), *Information and Software Technologies* (pp. 34-45). Communications in Computer and Information Science, vol 403. Heidelberg, Berlin: Springer.
- Raudys, A., & Pabarskaite, Ž. (2018). Optimising the Smoothness and Accuracy of Moving Average for Stock Price Data. *Technological and Economic Development of Economy*, 24(3), 984-1003.
- Saoud, L. S., & Al-Marzouqi, H. (2020). Metacognitive Sedenion-Valued Neural Network and its Learning Algorithm. *IEEE Access*, 8, 144823-144838.
- Shen, Z., Zhang, Y., Lu, J., Xu, J., & Xiao, G. (2020). A novel time series forecasting model with deep learning. *Neurocomputing*, 396, 302-313.
- Spinelli, A., & Pellino, G. (2020). COVID-19 pandemic: perspectives on an unfolding crisis. *British Journal of Surgery*, 107(7), 785-787.
- Svetunkov, I., & Petropoulos, F. (2018). Old dog, new tricks: a modelling view of simple moving averages. *International Journal of Production Research*, 56(18), 6034-6047.
- Usaratniwart, E., Sirisukprasert, S., Hatti, N., & Hagiwara, M. (2017). A case study in micro grid using adaptive enhanced linear exponential smoothing technique. *2017 8th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, Chonburi, Thailand, pp. 1-6.

Zakamulin, V. (2017). Types of Moving Averages. In: *Market Timing with Moving Averages. New Developments in Quantitative Trading and Investment* (pp. 23-52). Cham: Palgrave Macmillan.

Draft