American Journal of Public Health Research, 2021, Vol. 9, No. 6, 248-256
Available online at http://pubs.sciepub.com/ajphr/9/6/4
Published by Science and Education Publishing
DOI:10.12691/ajphr-9-6-4

A Comparative Study between ARIMA Model, Holt-Winters – No Seasonal and Fuzzy Time Series for New Cases of COVID-19 in Algeria

Abdelmounaim Hadjira[1,*], Hicham Salhi[2], Fadoua El Hafa[3]

[1]Department of Economics, Mohamed Boudiaf University, M'sila, Algeria
[2]Laboratory of Applied Research in Hydraulics, University of Mustapha Ben Boulaid-Batna 2, Algeria
[3]Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, 211166, China
*Corresponding author: Abdelmounaim.hadjira@univ-msila.dz

Received October 08, 2021; Revised November 10, 2021; Accepted November 19, 2021

Abstract **Background:** Coronavirus disease has become a worldwide threat affecting almost every country in the world. The spread of the virus is likely to continue unabated. The aim of this study is to compare between Autoregressive Integrated Moving Average (ARIMA) model, Fuzzy time series and Holt-Winters – No seasonal for forecasting the COVID-19 new cases in Algeria. **Methods:** Three different models to predict the number of Covid-19 new cases in Algeria were used. The number of new cases of COVID-19 in Algeria during the period from 24th February 2020 to 31th July 2021 was modeled according to ARIMA(4,1,2) model, Five based Fuzzy time series models including the Chen model, Heuristic Huareng model, Singh model, Abbasov-Manedova model and NFTS model, and Holt-Winters – No seasonal. **Results:** The predictive values were obtained from the 1st August 2021 to 31th December 2021. According to a set of criteria (*ME, MAE, MSE, RMSE, U*), we found that the FTNS model is the most accurate and best generating model for the values of the number of new cases of Covid-19. **Conclusion:** To the best of our knowledge, this is the first comparative study of three models of forecasting of Covid-19 new cases in Algeria. This study shows that ARIMA models with optimally selected covariates are useful tools for monitoring and predicting trends of COVID-19 cases in Algeria. Moreover, this forecast will help the Health authorities to be better prepared to fight the epidemic by engaging their healthcare facilities.

Keywords: *Covid-19, ARIMA, fuzzy time series model, Holt-Winter-non-seasonal, Algeria*

Cite This Article: Abdelmounaim Hadjira, Hicham Salhi, and Fadoua El Hafa, "A Comparative Study between ARIMA Model, Holt-Winters – No Seasonal and Fuzzy Time Series for New Cases of COVID-19 in Algeria." *American Journal of Public Health Research*, vol. 9, no. 6 (2021): 248-256. doi: 10.12691/ajphr-9-6-4.

1. Introduction

The world is facing a global health pandemic that health authorities worldwide have been unable to stop its spread. The 2019 novel coronavirus (2019-nCoV) was originally discovered on December in Wuhan, a city in the Hubei province in China [1]. The virus which is linked to the same family of viruses as Severe Acute Respiratory Syndrome (SARS) causes respiratory diseases which may lead to serious complications including death, has become a worldwide threat affecting almost every country in the world. According to the latest data provided by the world health organization (WHO), the total number of confirmed cases with Covid-19 virus has surpasses 219 million, and the number of deaths have reached 4.5 million. Moreover, these numbers continue to grow at a significant speed, indicating that the pandemic is far away from being over. Furthermore, mutant viral strains continue to emerge, and no public policies or treatment strategies seem to be sufficiently effective in blocking Covid-19 [2].

On February 25, 2020, Algeria laboratory-reported its first case of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), an Italian man who arrived on 17 February [3]. Over time, the number of confirmed cases of Covid-19 has increased, whether it is spread locally or imported cases from other countries. China is considered to be the most important economic trader in Africans countries including Algeria, which increases the probability of the virus spreading in Africa. A recent study have shown that Algeria and Egypt have moderate to high capacity to respond to outbreak, yet these two countries have the highest importation risk [3]. According to the Algerian government, on 02 August 2021, the number of infections was 170,000 and the death total was 4,189. Moreover, the number are at increase.

The mathematical and statistical models have been used in epidemiology in order to understand the dynamics of infectious diseases. Autoregressive Integral Moving Average model (ARIMA), namely the Box–Jenkins model, is the most common time series prediction model in the statistic model. In fact, several studies have attempted to track the trend of spread of Covid-19 virus and predict future path using statistical and mathematical methods. For instance, a previous research article have analyzed the time series data for the top five countries affected by Covid-19 namely US, Brazil, India, Russia and Spain, for forecasting the spread of the epidemic, using ARIMA models. The forecast accuracy were found within acceptable agreement predictions measured by MAD and MAPE [4].

Moreover, based on of the COVID-19 in Hubei the Chinese province, researcher have established the ARIMA model, then used these models to predict the development of the Italian epidemic in the next 10 days, even though the model fits well, yet is more suitable for short-term forecasting [5].

Several studies have used the ARIMA models to analyze the spread of Covid-19, using Japanese and South Korean data [6], also in Nigeria, using NCDC available data. In order to establish a suitable forecasting model, the daily data from February 27 to April 26, 2020 was collected, and an ARIMA model was constructed using R software. The model also carried out a 10-day forecast. The model showed that the spread of COVID-19 in Nigeria was on the rise during the selected time frame [7].

Relevant governments and health authorities can use these predictive models to have a better insight of the pandemic situation, and manage and strengthen their healthcare facilities for a better control and contain the outbreak in an efficient way.

Given the fact that the pandemic is ravaging life in Algeria, it becomes necessary to propose a predictive model for the spread of the pandemic.

A previous study have investigated the trend of spread of the virus in the most affected African countries: Algeria, Egypt, and South Africa. Two statistical approaches were used: a class of ARIMA model: Box-Jenkins and fuzzy time series to predict the incidence of Covid-19 cases. The results showed that trends in all three countries are positive, however, the fuzzy time series is better than the ARIMA model in forecasting. This study was the only one conducted in Algeria, however, the time period used is relatively short, only 32 days, and it was restrained on just two models [8].

The aim of this study, is to propose a predictive model for the spread of the SARS-CoV-19 virus resulting new Covid-19 cases in Algeria, using three different statistical and mathematical models, including: ARIMA, Holt-Winter models and fuzzy time series.

To the best of our knowledge, this is the first comparative study based on three different statistical models conducted in Algeria, and over a long time period of 17 months.

2. Data and Methodology

2.1. Data Source

Data was obtained from the reliable Our World in Data, an official website, which published a daily updated statistics of the coronavirus pandemic. This open access and open-source platform provide official daily and updated statistics for all countries of the world on a wide range of variables related to Coronavirus pandemic, such as: the number of accumulated cases, the number of new cases, the number of deaths, globally and by country [9]. For our study, we have extracted the number of confirmed new cases of COVID-19, the data obtained extends over a period of 17 months, from the period 24 February 2020 to 30 July 2021.

2.2. Methods

In our study, we used three different predictive methods to forecast the number of covid-19 new cases in Algeria. Box-Jenkins methodology that relies on ARIMA models, fuzzy time-series method including 5 different approaches: the Chen model, Heuristic-Huareng model, Singh model, Abbasov-Manedova model and NFTS model, and finally, the exponential smoothing models (Holt-Winters – No seasonal). The two most widely used approaches to time series forecasting are exponential smoothing and ARIMA models. ARIMA models try to characterize the autocorrelations in the data, whereas exponential smoothing methods were based on a description of trend and seasonality in the data [10], in fuzzy time series, forecasting several methods have been developed for constructing fuzzy relationships on linguistic-valued time-series data to obtain predictive values.

2.3. ARIMA Model - Box-Jenkins Methodology

In time series analysis, the Box Jenkins (ARIMA) model introduced by Box and Jenkins is one of the most widely used approaches [11]. Due to its generality, it can be applied to any series: stationary or not, with or without seasonal elements, Moreover, it can be implemented in a wide range of different programs [12].

ARMA model is a combination of autoregressive models $AR(p)$, which regresses on its own lagged terms and can be expressed as follows:

$$X_t = \varnothing_1 X_{t-1} + \varnothing_2 X_{t-2} + \dots\dots\dots + \varnothing_p X_{t-p} + \varepsilon_t \quad (1)$$

Where X_t is stationary, and $\varepsilon_t \sim wn(0, \sigma_\varepsilon^2)$, $\emptyset_i : (i = 1,2 \dots p)$ are model parametres.

Introducing the lag operator L , the equation (1) can be written as

$$\varnothing(L) X_t = \varepsilon_t$$

Where:

$$\varnothing(L) = 1 - \varnothing_1 L - \varnothing_2 L^2 - \dots\dots\dots - \varnothing_p L^p)$$

The second part of combination is Moving-average model $MA(q)$ builds a function of error terms of the past and can be expressed as follows:

$$X_t = \theta_1 \varepsilon_{t-1} + \dots\dots\dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Where $\varepsilon_t \sim wn(0, \sigma_\varepsilon^2)$ and $\theta_i : (i = 1,2 \dots q)$.

The equation (2) can be written as:

$$X_t = \left(1 + \theta_1 L + \theta_2 L^2 + \dots\dots\dots + \theta_q L^q\right) \varepsilon_t = \theta(L) \varepsilon_t$$

The $ARMA(p, q)$ model is defined as:

$$X_t = \varnothing_1 X_{t-1} + \varnothing_2 X_{t-2} + \ldots\ldots\ldots + \varnothing_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots\ldots\ldots + \theta_q \varepsilon_{t-q}$$

Or even more concisely as:

$$\varnothing(L) X_t = \theta(L)\varepsilon_t$$

If the series is non stationary, we generalize the ARMA model to become the ARIMA model by including integrated components.

According to Box and Jenkins (ARIMA) the steps of modeling a time series consists of three main components: Identification model, estimation model and diagnostic model [13].

Based on these main steps, we proposed a predictive model of the number of new cases of COVID-19 in Algeria which follows the steps below:

1. Check for the stationary of the time series through the graphic curve and the autocorrelation function correlogram, as well as carrying out the Dickey-Fuller test, Philips Peron test, KPSS test.
2. Identification of a tentative model by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF)
3. Estimating the various possible models using the method of maximum likelihood and choosing the best model based on Akaike criterion.
4. Estimating the appropriate model using the maximum likelihood method of estimation.
5. Diagnostic checking model.

2.4. Fuzzy Time Series

The second approach we will use in our comparative study is Fuzzy time series, described below:

Let U be the universe of discourse, $U = \{u1, u2, \ldots, um\}$.

A fuzzy set A of U is:

$$A = \left\{ \frac{\mu A(u1)}{u1}, \frac{\mu A(u2)}{u2}, \ldots, \frac{\mu A(um)}{um} \right\};$$

$\mu A(ui)$ is the grade of membership of ui in A, with: $0 \le \mu A(ui) \le 1$ and $i \in [0, m]$.

For the sake of accuracy, we will give some of the most used definitions [14,15,16,17]:

Definition 1. The first order relation defined by:

$$F(t) = F(t-1) oR(t, t-1).$$

If $F(t)$ is caused by $F(t-1)$ or $F(t-2)$, or... $F(t-m)$ with m>0.

Definition 2. The mth order relation defined by:

$$F(t) = \left(F(t-1) \times F(t-2) \times \ldots F(t-m)\right) oR(t, t-m).$$

If $F(t)$ is caused by $F(t-1)$ and $F(t-2)$ and… $F(t-m)$ with m>0.

Definition 3. Time variant fuzzy time series:

F(t) is called time variant fuzzy time series If the first and m^{th} order relation $R(t, t-1)$, or $R(t, t-m)$ of F(t) is independent of t.

$$R(t1, t1-1) = R(t2, t2-1), \; or \; R(t1, t1-1) = R(t2, t2-1).$$

2.5. Holt-Winters – No Seasonal

The third approach is the Holt-Winter method introduced by Holt in 1957 and Winter in 1960 [18,19], All exponential smoothing methods share the property of giving recent values relatively more weight in prediction than old observations [20]. Our study will be limited to the application of the Holt-Winter non seasonal method, which is similar to the double smoothing method that generates predictive values with a linear trend and without the seasonal component. This method use two smoothing constants α and β and three equation:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$$
$$F_{t+m} = L_t + b_t m$$

Where L_t and b_t denote an estimate of level and slope of series at time t, F_{t+m} Predictive value on the horizon m.

3. Results and Discussion

3.1. Box-Jenkins Methodology

The first step is visual examination of the data plot and the correlogram of the autocorrelation function (ACF), as it is shown in Figure (1,a) the series does not fluctuate around a fixed average and that the variance of the series does not appear to be constant.

The Figure (1,c,d) represents the correlogam of the covid-19 series, we can deduct that half of the autocorrelations coefficients are significantly different from zero and decrease exponentially. This is indicative of a non-stationary series. However, it is necessary to verify this assumption by applying statistical stationary tests such as: ADF test, Phillips – Perron test and KPSS test.

Table 1 represents the results of the stationary tests (ADF test, PP test, KPSS test) of the new cases series at the level and after the first differentiation, at the level, we notice that the p-value of ADF test and P-P test equal to 0.173 and 0.793 respectively, greater than 0.05 which means that the new cases series is not stationary, and also the results of KPSS test show that the calculated value is equal to 0.174 greater than the tabulated value (0.146) which confirms the results ADF and PP test. The new cases series is not stationary (has a unit root) and the best way to make it stationary is to differentiate it. Table 1 also shows that the p-value of ADF test and PP test of the differentiate series are < 0.001, <0.001 respectively, which indicates that the new cases series is stationary after the first differentiation, same result for the KPSS test.

Figure (1, b) shows the graph of the stationary series new cases after the first differentiation, through these stationary results we can conclude that the appropriate model is ARIMA (I = 1).

The second step of the Box-Jenkins methodology is to choose the appropriate model among the possible models by studying the plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF). Then choose the optimal model based on the Akaike criterion which is calculated as follows:

$$AIC = \log \hat{\sigma}^2_{\hat{\varepsilon}} + \frac{2(p+q)}{T}$$

The optimal model to fit the new cases series is the model which has the lowest Akaike information criterion. Relying on the auto-selection of model order based on Python library (pmdarima), we estimated 28 possible models, where we set max_p=7, max_q=7, Based on the minimum value of the Akaike criterion, we found that the appropriate model is ARIMA (4, 1, 2).

Table 2 represents the estimation results of the ARIMA model (4, 1, 2), the results shows that all the estimated parameters are significant at 1%, the p-value of all parameters are completely less than 0.01. The estimated model is of good quality if the calculated series follows the evolutions of the observed series as it is shown in Figure 2, the residuals between these values must therefore behave like white noise. The critical probability of the Ljung-Box statistic is always greater than 0.05 and all the terms of the correlogram are in the confidence interval (Figure 2). After estimating the model and verifying its validity, we predicted the number of cases infected with COVID-19 during the period: 08/01/2021 to 12/31/2021 with the confidence interval for the predicted values as shown in the Figure 3.

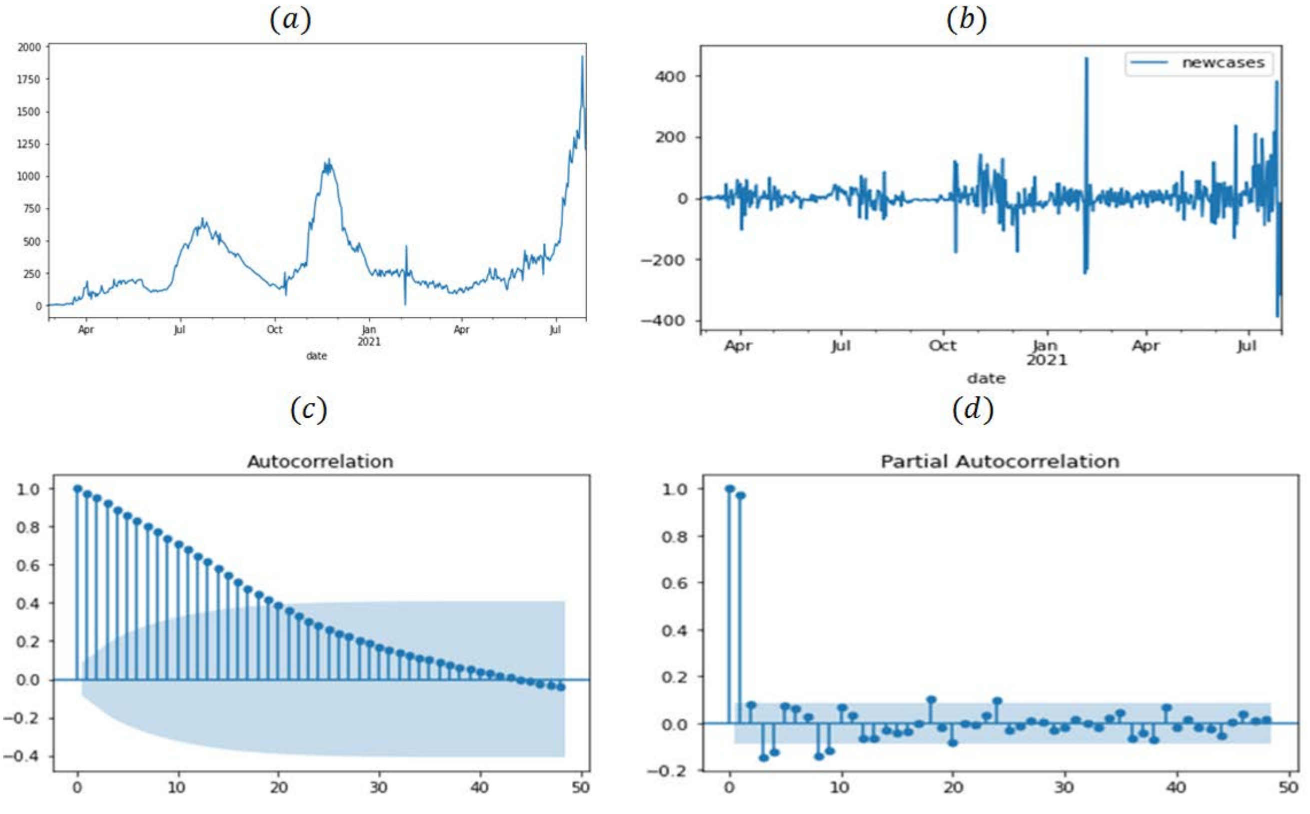

Figure 1. Series graph at and after differentiation, ACF and PACF correlogram

Table 1. ADF, PP and KPSS Test for Unit Root (with trend and intercept)

Variable	At level			First difference		
	ADF	PP	KPSS	ADF	PP	KPSS
New cases	-2.29	0.38	0.17	-4.60**	-29.93**	0.19
	(0.17)	(0.79)	[0.14]	(0.00)	(0.00)	0.46

** indicate significance at 1%. For ADF and PP; H_0 variable has a unit root and H_0 variable is stationary for KPSS test.

Table 2. Estimation ARIMA (4, 1, 2) model

$\boldsymbol{ARIMA}(\mathbf{4},\mathbf{1},\mathbf{2})$	Coefficient	Std.error	t-student	P-value
$\boldsymbol{AR}(\mathbf{1})$	1.2361	0.086	14.377	<0.001
$\boldsymbol{AR}(\mathbf{2})$	-0.1475	0.048	-3.091	0.002
$\boldsymbol{AR}(\mathbf{3})$	-0.3859	0.034	-11.333	<0.001
$\boldsymbol{AR}(\mathbf{4})$	0.2239	0.045	4.939	<0.001
$\boldsymbol{MA}(\mathbf{1})$	-1.6230	0.085	-19.032	<0.001
$\boldsymbol{MA}(\mathbf{2})$	0.7373	0.074	9.994	<0.001
$\boldsymbol{sigma2}$	2504.3511	67.600	37.047	<0.001
Ljung-Box (L1) (Q)	0.00	Proba	0.96	
Heteroskedasticity (H):	6.91	Proba	0.00	

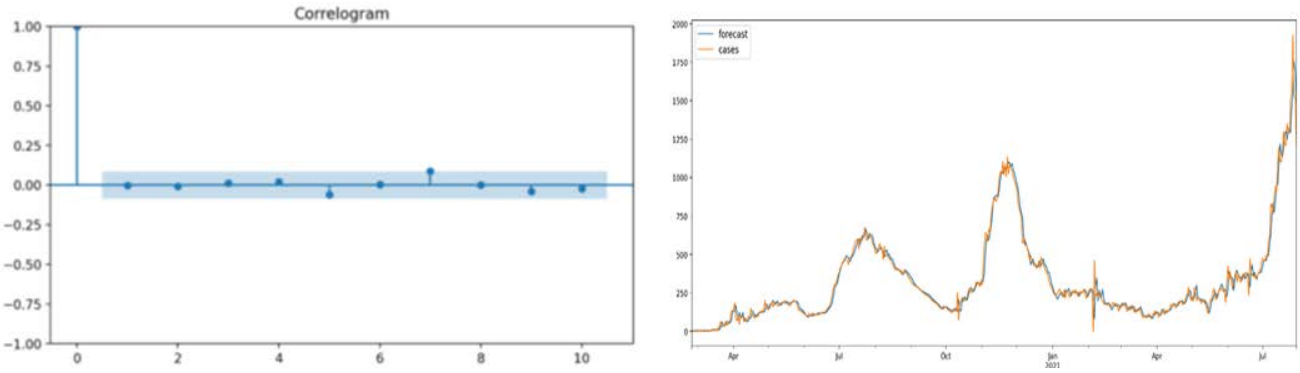

Figure 2. Correlogram of residual, Actual and Forecasting value

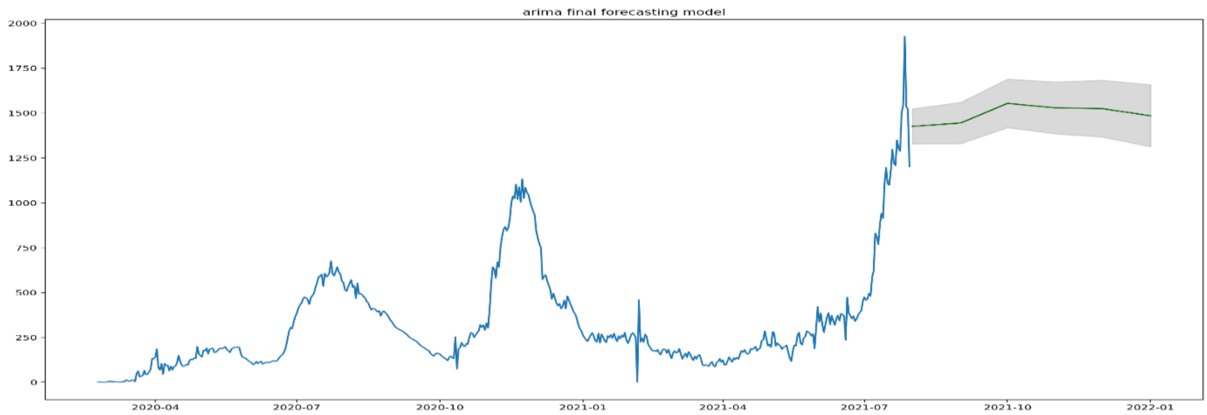

Figure 3. Covid-19 new cases forecast plot

3.2. Fuzzy Time Series

The following steps have been performed to estimate new COVID cases in Algeria:

Step 1. Define universe of discourse U as

$$U = [U_{\min}, U_{\max}]$$

Where U_{min} and U_{max} are values of real time new case data respectively.

$$U = [0, 1927]$$

Step 2. Partitioning U into equal length of sub intervals based on formula statistic

$$K = 1 + 3.32 * \log(T)$$

with T: number of observations T=523.

Table 3. Fuzzy sets determination

	Set	Dow	Up	Mid	Num
1	u_1	0.0	192.7	96.35	220
2	u_2	192.7	385.4	289.05	161
3	u_3	385.4	578.1	481.75	75
4	u_4	578.1	770.8	674.45	26
5	u_5	770.8	963.5	867.15	16
6	u_6	963.5	1156.2	1059.8	17
7	u_7	1156.2	1348.9	1252.55	8
8	u_8	1348.9	1541.6	1445.25	4
9	u_9	1541.6	1734.3	1637.95	1
10	u_{10}	1734.3	1927.00	1830.65	1

We find $K = 10$. Table 3 shows the sub intervals of U.

Step 3. Defining each fuzzy setA_i, whith 1< i <10.

$$A_1 = 1/U_1 + 0.5/U_2 + 0/U_3 + 0/U_4 + 0/U_5 + 0/U_6 + 0/U_7 + 0/U_5 + 0/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_2 = 0.5/U_1 + 1/U_2 + 0.5/U_3 + 0/U_4 + 0/U_5 + 0/U_6 + 0/U_7 + 0/U_5 + 0/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_3 = 0/U_1 + 0.5/U_2 + 1/U_3 + 0.5/U_4 + 0/U_5 + 0/U_6 + 0/U_7 + 0/U_5 + 0/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_4 = 0/U_1 + 0/U_2 + 0.5/U_3 + 1/U_4 + 0.5/U_5 + 0/U_6 + 0/U_7 + 0/U_5 + 0/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_5 = 0/U_1 + 0/U_2 + 0/U_3 + 0.5/U_4 + 1/U_5 + 0.5/U_6 + 0/U_7 + 0/U_5 + 0/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_6 = 0/U_1 + 0/U_2 + 0/U_3 + 0/U_4 + 0.5/U_5 + 1/U_6 + 0.5/U_7 + 0/U_5 + 0/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_7 = 0/U_1 + 0/U_2 + 0/U_3 + 0/U_4 + 0/U_5 + 0.5/U_6 + 1/U_7 + 0.5/U_5 + 0/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_8 = 0/U_1 + 0\ /U_2 + 0/U_3 + 0/U_4 + 0/U_5 + 0/U_6 + 0.5/U_7 + 1/U_5 + 0.5/U_8 + 0/U_9 + 0/U_{10}.$$

$$A_9 = 0/U_1 + 0\ /U_2 + 0/U_3 + 0/U_4 + 0/U_5 + 0/U_6 + 0/U_7 + 0.5/U_5 + 1/U_8 + 0.5/U_9 + 0/U_{10}.$$

$$A_{10} = 0/U_1 + 0\ /U_2 + 0/U_3 + 0/U_4 + 0/U_5 + 0/U_6 + 0/U_7 + 0/U_5 + 0/U_8 + 0.5/U_9 + 1/U_{10}.$$

Step 4. Fuzzy logical relationships of 10 groups, obtain a total of 250 logical relations

$$"A1->A1,A2,A3"\quad "A2->A1,A2,A3"$$
$$"A3->A2,A3,A4"\quad "A4->A3,A4,A5"$$
$$"A5->A4,A5,A6"\quad "A6->A5,A6,A7"$$
$$"A7->A1,A6,A7,A8"\quad "A8->A7,A8,A9"$$
$$"A9->A10"\quad "A10->A8"$$

Step 5: Using the R Package "AnalyseTS", we have estimated the predictive values of a new cases series using five models in fuzzy time series analysis, the actual and forecast values of new cases covid-19 are showing in Figure 4.

- Chen (1996) model [19] (Figure 4(a)).
- Singh (2008) model [20] (Figure 4(b)).
- Heuristic Huareng (2001) model [21] (Figure 4(c)).
- Abbasov-Manedova (2010) model [22] (Figure 4(d)).
- New fuzzy time series (NFTS) model [23] (Figure 4(e)).

3.3. Holt-Winters – No Seasonal

We used the Eviews 12 program to perform the prediction using the Holt –Winters-No seasonal approach, and obtain the predictive values of the number of cases of Covid-19 during the time period from 08/01/2021 until 12/31/2021. We found that the estimated value of $\alpha = 0.61, \beta = 0.10$. Figure 5 shows actual and forecast values for the number of new COVID-19 cases.

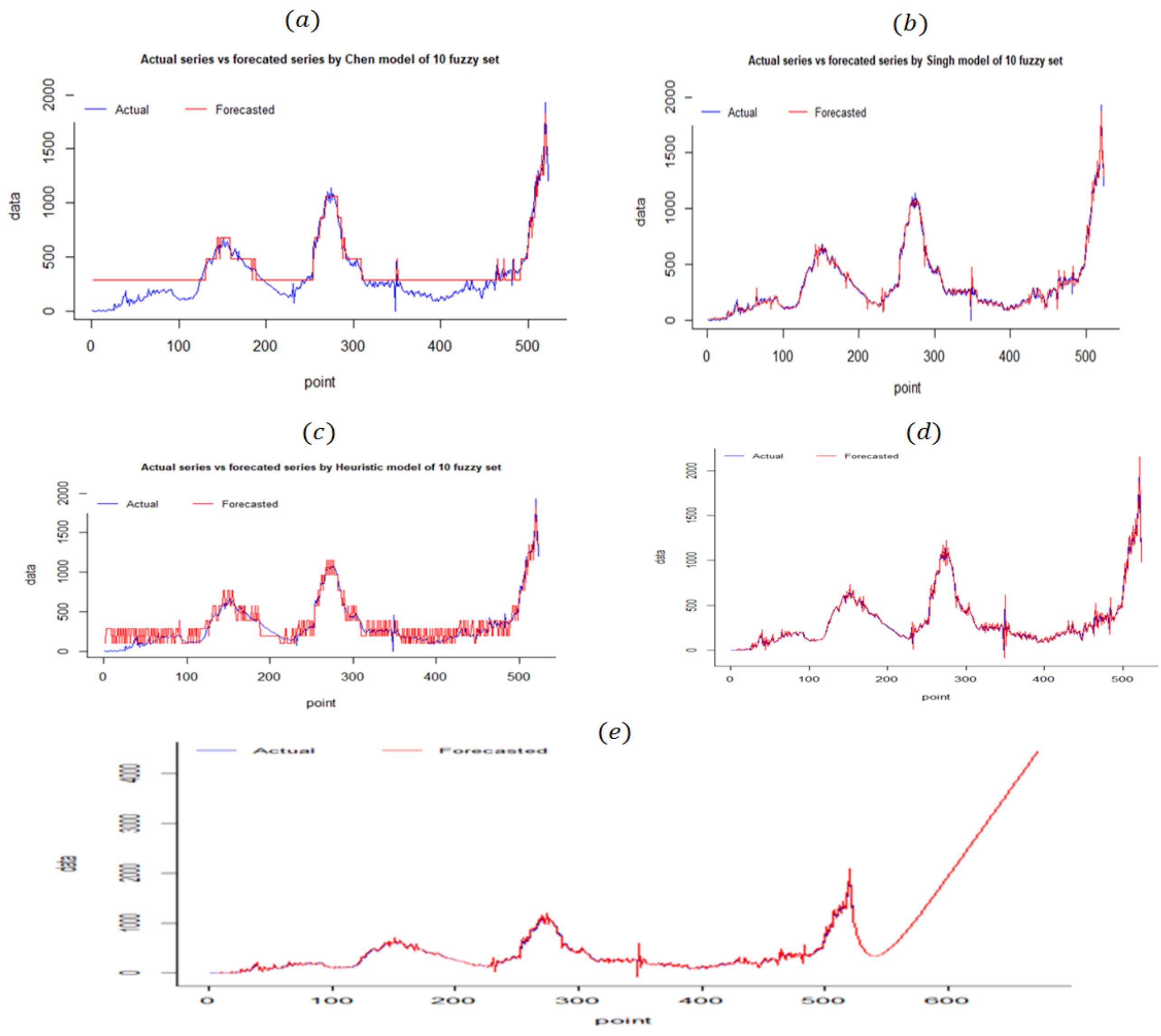

Figure 4. Actual and forecast values of 5 different models of FTS

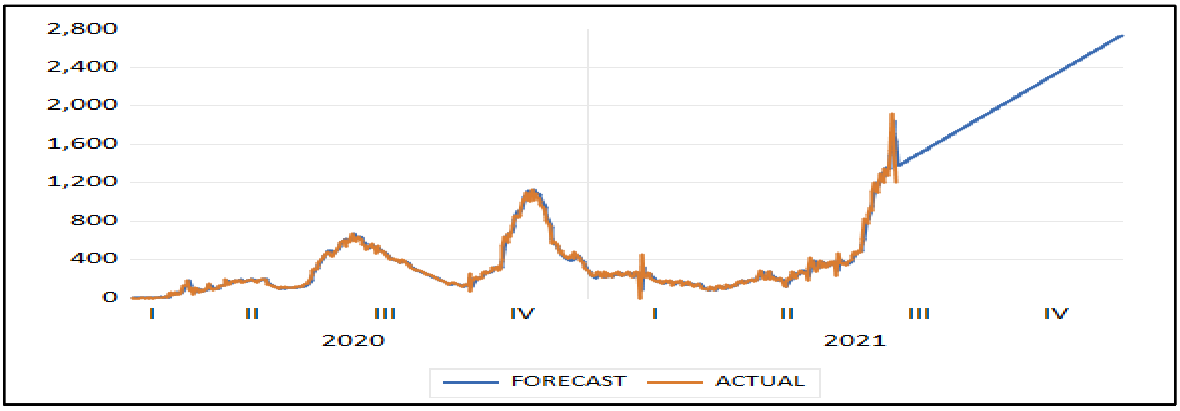

Figure 5. Actual and forecast value of Covid-19 new cases

A comparison of different forecasting methods

According to our results, and in order to choose the best model that gives predictive values for COVID-19 cases in Algeria, we calculated five selection criteria: ME, MAE, MSE, RMSE and U.

$$ME = \frac{\sum(forecast - actual)}{T}$$

$$MAE = \frac{\sum|forecast - actual|}{T}$$

$$MSE = \frac{\sum(forecast - actual)^2}{T}$$

$$RMSE = \sqrt{\frac{\sum(forecast - actual)^2}{T}}$$

$$U = \frac{\sqrt{\sum_{t=1}^{n-1}\left(\frac{forecast_{i+1} - actual_i}{actual_i}\right)^2}}{\sqrt{\sum_{t=1}^{n-1}\left(\frac{actual_{i+1} - actual_i}{actual_i}\right)^2}}$$

Table 4 shows the values of the statistical criteria ME, MAE, MSE, $RMSE$, U for the $ARIMA(4,1,2)$ model, the Holt-Winter model, and the five fuzzy time-series models. First, comparing with both $ARIMA(4,1,2)$ and Holt winter models the fuzzy time series models show the minimum values of the criteria: $MAE, MSE, RMSE, U$. As Table 4 shows: the $RMSE = 25.239$ for the NFTS model, while for the $ARIMA(4,1,2)$ model the $RMSE\ is\ 474.605$. Moreover, we can deducted based on the results of these selection criteria that the fuzzy time-series models outperform the ARIMA and Holt-Winter models in predicting the number of new cases of COVID-19 in Algeria.

However, we used five different models of the fuzzy time-series approach. We relied on the same selection criteria to choose the most accurate fuzzy time series model.

According to the criteria MAE, MSE, RMSE, U, it is clear that the NFTS model has the lowest values compared to the Chen model, Heuristic-Huareng model, Singh model, Abbasov-Manedova model. Based on these results, we conclude that the NFTAS model is the most accurate model that provides predictive values for the number of Covid-19 cases in Algeria. Table 5 shows the predictive values of new cases of COVID-19 during the time period from 1st August 2021 to 31th December 2021.

Table 4. Compare models by criteria: MAE, MSE, $RMSE$, U

Models	ME	MAE	MSE	RMSE	U
ARIMA(4,1,2)	-262.41	274.76	225249.9	474.60	-
Holt-Winters	-672,72	396,09	2734.98	52.29	-
Chen (1996)	-25.73	59.20	7348.15	85.72	1.55
Heuristic Huareng (2001)	-33.80	89.51	12422.83	111.45	2.01
Singh (2008)	-0.22	24.63	1130.76	33.62	0.60
Abbasov-Manedova (2010)	-5.82	37.41	5001.89	70.72	1.27
NFTS	-7.84	15.04	637.04	25.23	0.45

Table 5. Predictive values of COVID-19 new cases from 01/08/2021 to 31/12/2021

Date	Forecast	Diff.forecast	Date	Forecast	Diff.forecast	Date	Forecast	Diff.forecast
8/1/2021	957	-147,805	10/1/2021	1394	33,02	12/1/2021	3479	34,48
8/2/2021	866	-98,46	10/2/2021	1427	33,12	12/2/2021	3514	34,48
8/3/2021	791	-90,41	10/3/2021	1461	33,22	12/3/2021	3548	34,48
8/4/2021	715	-75,29	10/4/2021	1494	33,31	12/4/2021	3583	34,48
8/5/2021	649	-76,19	10/5/2021	1527	33,39	12/5/2021	3617	34,48
8/6/2021	592	-66,16	10/6/2021	1561	33,47	12/6/2021	3686	34,48

Date	Forecast	Diff.forecast	Date	Forecast	Diff.forecast	Date	Forecast	Diff.forecast
8/7/2021	543	-56,51	10/7/2021	1594	33,54	12/7/2021	3720	34,48
8/8/2021	500	-49,17	10/8/2021	1628	33,61	12/8/2021	3755	34,49
8/9/2021	463	-42,58	10/9/2021	1662	33,67	12/9/2021	3790	34,49
8/10/2021	432	-36,99	10/10/2021	1695	33,73	12/10/2021	3824	34,49
8/11/2021	406	-31,21	10/11/2021	1729	33,78	12/11/2021	3859	34,49
8/12/2021	385	-25,82	10/12/2021	1763	33,83	12/12/2021	3893	34,49
8/13/2021	369	-20,99	10/13/2021	1797	33,88	12/13/2021	3928	34,49
8/14/2021	356	-16,62	10/14/2021	1831	33,92	12/14/2021	3962	34,49
8/15/2021	347	-12,65	10/15/2021	1865	33,96	12/15/2021	3997	34,49
8/16/2021	341	-9	10/16/2021	1899	34	12/16/2021	4031	34,49
8/17/2021	339	-5,66	10/17/2021	1933	34,06	12/17/2021	4066	34,49
8/18/2021	342	-2,63	10/18/2021	1967	34,09	12/18/2021	4100	34,49
8/19/2021	347	0,12	10/19/2021	2001	34,12	12/19/2021	4135	34,49
8/20/2021	354	2,66	10/20/2021	2035	34,15	12/20/2021	4169	34,49
8/21/2021	363	5	10/21/2021	2069	34,17	12/21/2021	4204	34,49
8/22/2021	374	7,15	10/22/2021	2103	34,2	12/22/2021	4238	34,49
8/23/2021	386	9,13	10/23/2021	1238	34,22	12/23/2021	4273	34,49
8/24/2021	401	10,96	10/24/2021	2172	34,24	12/24/2021	4307	34,49
8/25/2021	416	12,65	10/25/2021	2206	34,25	12/25/2021	4342	34,49
8/26/2021	433	14,21	10/26/2021	2240	34,27	12/26/2021	4376	34,49
8/27/2021	452	15,66	10/27/2021	2275	34,29	12/27/2021	4411	34,49
8/28/2021	471	16,99	10/28/2021	2309	34,3	12/28/2021	4410	34,49
8/29/2021	491	18,23	10/29/2021	2343	34,32	12/29/2021	4445	34,49
8/30/2021	513	19,38	10/30/2021	2378	34,32	12/30/2021	4480	34.49
8/31/2021	535	20,45	10/31/2021	2412	4,33	12/31/2021	4514	34.49
9/1/2021	558	21,44	11/1/2021	2446	34,33			
9/2/2021	582	22,36	11/2/2021	2481	34,34			
9/3/2021	607	23,21	11/3/2021	2515	34,35			
9/4/2021	633	24	11/4/2021	2549	34,36			
9/5/2021	659	24,75	11/5/2021	2584	34,37			
9/6/2021	633	25,42	11/6/2021	2618	34,38			
9/7/2021	659	26,06	11/7/2021	2653	34,39			
9/8/2021	685	26,65	11/8/2021	2687	34,39			
9/9/2021	713	27,2	11/9/2021	2721	34,4			
9/10/2021	740	27,71	11/10/2021	2756	34,41			
9/11/2021	768	28,19	11/11/2021	2790	34,41			
9/12/2021	797	28,63	11/12/2021	2825	34,42			
9/13/2021	826	29,04	11/13/2021	2859	34,42			
9/14/2021	856	29,42	11/14/2021	2893	34,43			
9/15/2021	885	29,78	11/15/2021	2928	34,43			
9/16/2021	915	30,11	11/16/2021	2962	34,44			
9/17/2021	946	30,41	11/17/2021	2997	34,44			
9/18/2021	977	30,7	11/18/2021	3031	34,45			
9/19/2021	1008	30,97	11/19/2021	3066	34,45			
9/20/2021	1039	31,21	11/20/2021	3100	34,45			
9/21/2021	1070	31,44	11/21/2021	3135	34,46			
9/22/2021	1102	31,66	11/22/2021	3169	34,46			
9/23/2021	1134	31,85	11/23/2021	3204	34,46			
9/24/2021	1166	32,04	11/24/2021	3238	34,47			
9/25/2021	1198	32,21	11/25/2021	3273	34,47			
9/26/2021	1230	32,37	11/26/2021	3307	34,47			
9/27/2021	1263	32,52	11/27/2021	3341	34,47			
9/28/2021	1296	32,66	11/28/2021	3376	34,47			
9/29/2021	1328	32,79	11/29/2021	3410	34,48			
9/30/2021	1361	32,91	11/30/2021	3445	34,48			

4. Conclusion

Finding the pattern or model that generates values for the number of new cases of Covid-19 is very crucial. In our study, we tested and compared the predictive ability of different models using data of Covid-19 infection cases from the period 24 February 2020 to 30 July 2021. The most important results were as follows:

- Following the Box Jenkins methodology, which relies on ARIMA models, our study concluded that the best model that gives predictive values is the $ARIMA(4,1,2)$ model.
- Using the exponential smoothing models and estimating the Holt-Winters – No seasonal model, through which we compare the current and forecast values and generate predictive values during the period 01 August 2021 to 31 December 2021.
- Relying on a set of fuzzy time series models (the Chen model, Heuristic Huareng model, Singh model, Abbasov-Manedova model, and NFTS model) we generated the predictive values and compared them with the actual values.
- Relying on a set of statistical criteria (ME, MAE, MSE, $RMSE$, U), we made a comparison between all these models, and the results revealed that the best model is the NFTS model, then we presented the predictive values during the period 01 Augest 2021 until 31 December 2021.

The main objective of this study is to provide predictive values for the expected number of COVID-19 cases to take the necessary precautions measures. According to our results, the most accurate forecast model shows that the number of cases will increase in the upcoming days, therefore the health authorities should maintain restrictions on gatherings, as well as social distancing and wearing a mask.

References

[1] Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med.2020; 382: 1199-1207.

[2] Aleem A, Akbar Samad AB, Slenker AK. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). StatPearls. Treasure Island (FL): StatPearls Publishing LLC.; 2021.

[3] GILBERT, Marius, PULLANO, Giulia, PINOTTI, Francesco, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. The Lancet, 2020, vol. 395, no 10227, p. 871-877.

[4] SAHAI, Alok Kumar, RATH, Namita, SOOD, Vishal, et al. ARIMA modelling & forecasting of COVID-19 in top five affected countries. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020, vol. 14, no 5, p. 1419-1427.

[5] YANG, Qiuying, WANG, Jie, MA, Hongli, et al. Research on COVID-19 based on ARIMA modelΔ—Taking Hubei, China as an example to see the epidemic in Italy. Journal of Infection and Public Health, 2020, vol. 13, no 10, p. 1415-1418.

[6] DUAN, Xingde et ZHANG, Xiaolei. ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data. Data in brief, 2020, vol. 31, p. 105779.

[7] IBRAHIM, Rauf Rauf et OLADIPO, OLUWAKEMI HANNAH. Forecasting the spread of COVID-19 in Nigeria using Box-Jenkins modeling procedure. medRxiv, 2020.

[8] FATIH, Chellai, HAMIMES, Ahmed, et MISHRA, Pradeep. A note on Covid-19 Statistics, Strange trend and Forecasting of Total Cases in the most Infected African Countries: An ARIMA and Fuzzy Time Series Approaches. African Journal of Applied Statistics, 2020, vol. 2, no 2, p. 961-975.

[9] Roser, Max, et al. “Coronavirus pandemic (COVID-19).” Our world in data (2020). https://ourworldindata.org/coronavirus.

[10] Hyndman, Rob J., and George Athanasopoulos. Forecasting: principles and practice. OTexts, (2014).

[11] Box, G.E.P. and G.M Jenkins. Time series Analysis Forecasting and Control, Holden-day, San Francisco. (1970).

[12] Maddala, Gangadharrao S., and In-Moo Kim. “Unit roots, cointegration, and structural change.” (1998).

[13] Box, George EP, et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

[14] Song, Q. and B.S. Chissom (1993b). Fuzzy time series and its models. Fuzzy Sets and Systems, 54, 269-277.

[15] RANA, A. K. Study on Fuzzy Time Invariant Series Models for Crop Production Forecasting. International Journal of Scientific Research and Reviews, 2019, vol. 8, no 2, p. 3729-3741.

[16] RAJARATHINAM, A. et THIRUNAVUKKARASU, M. Fuzzy Time Series Modeling for Paddy (Oryza sativa L.) Crop Production. 2013.

[17] VOVAN, Tai. An improved fuzzy time series forecasting model using variations of data. Fuzzy Optimization and Decision Making, 2019, vol. 18, no 2, p. 151-173.

[18] Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA.

[19] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324-342.

[20] Wheelwright, Steven, Spyros Makridakis, and Rob J. Hyndman. Forecasting: methods and applications. John Wiley & Sons, 1998.

[21] CHEN, Shyi-Ming. Forecasting enrollments based on fuzzy time series. Fuzzy sets and systems, 1996, vol. 81, no 3, p. 311-319.

[22] Singh, Shiva Raj. “A computational method of forecasting based on fuzzy time series.” Mathematics and Computers in Simulation 79.3 (2008): 539-554.

[23] Huarng, H., 2001. Huarng models of fuzzy time series for forecasting. Fuzzy Sets and Systems. 123: 369-386.

[24] Abbasov, A.M. and Mamedova, M.H., 2003. Application of fuzzy time series to population forecasting, Proceedings of 8th Symposion on Information Technology in Urban and Spatial Planning, Vienna University of Technology, February 25-March1, 545-552.

[25] Chen, S.M. and Hsu, C.C., 2004. A New method to forecast enrollments using fuzzy time series. International Journal of Applied Science and Engineering, 12: 234-244.