

Stochastic modelling for predicting COVID-19 prevalence in East Africa Countries



Rediat Takele

Assistant Professor in Bio-Statistics, Jigjiga University, Department of Statistics, Ethiopia

ARTICLE INFO

Article history:

Received 14 May 2020

Received in revised form 23 July 2020

Accepted 7 August 2020

Available online 18 August 2020

Handling Editor: Dr. J Wu

Keywords:

Corona virus case

COVID19

Pandemic

Prediction

Prevalence

ARIMA model

ABSTRACT

Coronavirus (COVID-19) has continued to be a global threat to public health. As the matter of fact, it needs unreserved effort to monitor the prevalence of the virus. However, applying an effective prediction of the prevalence is thought to be the fundamental requirement to effectively control the spreading rate. Time series models have extensively been considered as the convenient methods to predict the prevalence or spreading rate of the disease. This study, therefore, aimed to apply the Autoregressive Integrated Moving Average (ARIMA) modeling approach for projecting coronavirus (COVID-19) prevalence patterns in East Africa Countries, mainly Ethiopia, Djibouti, Sudan and Somalia.

The data for the study were obtained from the reports of confirmed COVID-19 cases by the official website of Johns Hopkins University from 13th March, 2020 to 30th June, 2020. The results of the study, then, showed that in the coming four month, the number of COVID-19 positive people in Ethiopia may reach up to 56,610 from 5,846 on June 30, 2020 in average-rate scenario. However, in worst case scenario forecast, the model showed that the cases will be around 84,497. The analysis further depicted that with average interventions and control scenario, cumulative number of infected persons in Djibouti, Somalia and Sudan will increase from 4,656, 2,904 and 9,258 respectively at the end of June to 8,336, 3,961 and 21,388, which is by the end of October, 2020, after four-months. But, with insufficient intervention, the number of infected persons may grow quickly and reach up to 14,072, 10,037 and 38,174 in Djibouti, Somalia and Sudan respectively. Generally, the extent of the coronavirus spreading was increased from time to time in the past four month, until 30 th June, 2020, and it is expected to continue quicker than before for the coming 4-month, until the end of October, 2020, in Ethiopia, Djibouti, Somalia, and Sudan and more rapidly than before in Sudan and Ethiopia, while the peak will remain unknown yet. Therefore, an effective implementation of the preventive measures and a rigorous compliance by avoiding negligence with the rules such as prohibiting public gatherings, travel restrictions, personal protection measures, and social distancing may alleviate the spreading rates of the virus, particularly, Sudan and Ethiopia. Moreover, more efforts should be exerted on Ethiopian side to control the population movement across all the border areas and to strengthen border quarantining. Further, through updating more new data with continuous reconsideration of predictive model, provide useful and more precise prediction. Applying, ARIMAX-Transfer Function model in region-wise by take in to consideration of climatic data like temperature and humidity in each countries looking spatial pattern for reliable measure of COVID-19 prevalence.

E-mail address: redtakele@gmail.com.

Peer review under responsibility of KeAi Communications Co., Ltd.

<https://doi.org/10.1016/j.idm.2020.08.005>

2468-0427/© 2020 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The first case of coronavirus (COVID-19) outbreak was reported on December 31, 2019 in Wuhan, the central province of China, and then, it has been conveyed the global pandemic owning severe new type of threat to human health and life in a continues method. According to [World Health Organization \(2020\)](#), more than 15 million COVID-19 cases, 600,000 deaths, and 9 million recoveries have been reported during the study period (<https://www.who.int/>, [John Hopkins University, 2020](#)).

When it comes to Africa, most of the cases were imported from European countries (Gilbert et al., 2020 cited also in Zebin Z., et al., 2020). Though the situation has been alleviated in China, it has been worthy increasing in Europe and America. It, recently, has also been worsening situation in Africa. Reportedly, there have been more than 750,000 confirmed cases, 15,000 deaths and 450,000 recoveries ([John HopkinsUniversity \(2020\)](#)). Particularly, South Africa, Egypt and Algeria are the most affected countries. Likewise, the pandemic has been reported as an exponentially spreading in East African countries such as Ethiopia, Djibouti, Somalia, Kenya and Sudan. There have been various factors that would facilitate the rapid spreading of the virus in those countries. To mention some, poor social-economic condition of the people, scarce medical supplies, poor medical conditions and low virus testing efficiency are the fundamental factors that would facilitate the rapid transmission of the pandemic.

However, the governments of East African countries have been trying to implement various precautionary measures such as developing public awareness about general behavior of the virus and its prevention mechanisms like sanitization of streets and markets, quarantine of suspected and infected cases, lockdown of the schools, churches, and selected services at different scales and even, declaring state of emergency. Specifically, the Ethiopian government has implemented precautionary measures immediately after the virus has been reported the world pandemic. However, the transmission has remained rapid yet because most of the people have been negligence in some of preventive measures such as social distancing and mass gathering. On the other, population movement across the border from the neighboring countries like Sudan, Djibouti, Kenya and Somalia has a lion share for the recent trend of rapid transmission of coronavirus in Ethiopia. Consequently, COVID-19 transmission would get serious in Ethiopia particularly and East Africa generally, and its future trend is also expected to depend directly or indirectly on the transmission of COVID-19 inside and outside of all those East African countries.

For that reason, an access to the valid epidemiological information on transmission dynamics, severity, susceptibility and the effects of control measures has the fundamental significance in dealing with the pandemic, Corona Virus or COVID-19. Accordingly, analyzing the trends and forecasting future indices of the virus in East African countries must be conducted in order to put in a place an effective controlling strategies. While the transmission mechanisms of the virus is not completely known, the number of confirmed cases are drastically increasing, and the effects of containment should be evaluated on empirical data where quantitative analysis is more relevant.

Indeed, many countries in the world so far have applied data-driven statistical models like Autoregressive Integrated Moving Average (ARIMA) for prediction of future trends of the infectious disease such as Dengue fever ([Luz et al., 2008](#) and [Wongkoon et al., 2012](#)), Hemorrhagic fever with renal syndrome ([Liu et al., 2011](#)) and Tuberculosis ([Rios et al., 2000](#)), and have ascertained the effectiveness of the model. Moreover, different statistical models have been proposed by different authors worldwide, to predict COVID-19 prevalence in different countries based on spread pattern of the disease. This includes models for China ([Abenvenuto et al., 2020](#), [Anastassopoulou et al., 2020](#), [Li et al., 2020](#), [Liu, Beeler, & Chakrabarty, 2020](#), [Liu, Magal, Seydi, & Webb, 2020](#), [Roosa et al., 2020](#), [Fanelli & Piazza, 2020](#), [Hu et al., 2020](#), , [Liu, Magal, Seydi, & Webb, 2020](#), [Wu et al., 2020](#)), Italy ([Fanelli & Piazza, 2020](#); [Grasselli, Pesenti, & Cecconi, 2020](#); [Russo et al., 2020](#); [Jia et al., 2020](#)), France ([Fanelli & Piazza, 2020](#); [Massonnaud et al., 2020](#)), USA ([Liu, Beeler & Chakrabarty, 2020](#); [Lover & McAndrew, 2020](#); [Wise et al., 2020](#)), South Korea ([Zhan et al., 2020](#); [Kim, 2020](#)), India ([Gupta & Pal, 2020](#)) and Kenya ([Yoner et al., 2020](#)). Generally, time series forecasting models like ARIMA is the standard technique which gives a decent predictions and forecasts on time series data in quick time. This technique of analysis has been widely applied, for its reliability and quick implementation by various stakeholders.

As to the knowledge of the researcher, however, no enough research have been done in modeling the prevalence of COVID-19 transmission in the East African countries, although the infection of the virus, backed with low-socio-economic levels, is so dispiriting and poses more difficulties to prevention and control. Therefore, the aim of this research is to apply the Autoregressive Integrated Moving Average (ARIMA) modeling approach for projecting coronavirus (COVID-19) prevalence in selected East African countries to;

- enable the public health institutions to conduct reliable daily forecast and to come up with suitable intervention strategies
- provide instantaneous and long-term prevalence trajectory of the disease
- contribute to the body of knowledge in epidemiologic study method to forecast

2. Material and method

2.1. Study area description

The study was conducted in East African countries, mainly, focused on Ethiopia, Djibouti, Somalia and Sudan, where the pandemic rapidly transmission and affecting. According to UN population estimates, total population of the countries were 114,963,583 peoples of Ethiopia, 988,002 of Djibouti, 15,893, 219 of Somalia and 43, 849,269 of Sudan. Ethiopia is located in the North Eastern part of the African continent. It is bounded by Sudan on the west, Eritrea and Djibouti on the northeast, Somalia on the east and southeast, and Kenya on the south. Ethiopia lies between the Equator and Tropic of Cancer, between the 3° N and 15° N Latitude or 33° E and 48° E Longitude. The average seasonal temperatures are Low of 10 °C and High of 39 °C. Sudan, Djibouti and Somalia also is located in East Africa. The climate of Djibouti is tropical desert on the coast and in the north, while it is semi-desert in the central-southern highlands. Somalia has a tropical but not torrid climate, and there is little seasonal change in temperature. In the low areas, the mean temperature ranges from about 24 °C to 31 °C.

2.2. Study data description

The daily time series data of Nobel Coronavirus (COVID-19) of Ethiopia, Sudan, Djibouti, and Somalia from 13 March 2020 to June 30, 2020 were collected from the official website of [Johns Hopkins university \(2020\): https://github.com/CSSEGISandData/COVID-19](https://github.com/CSSEGISandData/COVID-19). To build a time-series database Ms. Excel 2019 was used. And R-4.0.0 statistical software were applied to perform statistical data analysis on the confirmed COVID-19 case datasets.

2.3. Method of data analysis

Time series analysis aims to reveal reliable and meaningful statistics and use this knowledge to predict future values of the series (Elevli et al., 2016; He and Tao, 2018; Benvenuto et al., 2020). Time series models attempt to forecast the future values by analyzing the past and current. It considers historical data and attempts to derive some process which will explain those occurrences and predict future values. The special feature of time series analysis is the fact that successive observations are usually not independent and that the analysis must take techniques to identify the patterns which typically exist in the series. Based on this fact, among time series techniques, this study applied ARIMA model in order to assess the future trend scenario of coronavirus (COVID-19) in Selected East African countries.

2.3.1. Autoregressive Integrated Moving Average (ARIMA) modeling

It is the most frequently applicable time series models as it takes into account changing trends, periodic changes and random disturbances in the time series. It also suitable for all kinds of data, including non-stationary data, which is if there is no systematic change in mean (no trend), no systematic change in variance and periodic variations have removed (Shumway & Stoffer, 2010). In practice, most of the time series are non-stationary and Box and Jenkins (1976) recommends removing any non-stationary sources of variation in time series data. In most case, common technique for achieving stationary is to apply regular differencing and log transformation to the original time series (x_t). If differencing a series d times makes it into a stationary and then series said to follows an autoregressive integrated moving average process, denoted by ARIMA (p, d, q) and can be written as:

$$\phi(\beta)\nabla^d x_t = \theta(\beta)w_t \quad (1)$$

where:

- Autoregressive operator can be expressed as: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$
- Moving average operator ($\theta(B)$): $\theta(B) = 1 + \theta_1 B + \theta_2 B + \dots + \theta_q B$
- Differencing operator (∇^d) = $(1 - \beta)^d$, it is the expression of d th consecutive differencing so as to make series stationary (Vandale, 1983).
- w_t is a Gaussian white noise series with mean zero and variance (σ_w^2).

2.3.1.1. Testing for stationary. Before developing a Box-Jenkins modeling process, it is important to check whether the data under study meets basic assumptions such as series stationary. A time series is considered as stationary if its statistical properties such as mean, variance constant over time (Shumway & Stoffer, 2010). Many testing procedures for stationary are proposed in the literature. In this study, correlogram and Augmented Dickey-Fuller (ADF) test were applied for testing whether the series is stationary.

The correlogram test: It is one way to characterize a series with respect to its dependence over time. It commonly known as sample autocorrelation function (ACF), which is plot of sample ACF coefficient against observation difference in time (lag). As state in Wei (2006) if the sample ACF decays very slowly in non-seasonal and seasonal lags, it indicates that differencing is needed and an implication for series non-stationary.

The Augmented Dickey-Fuller (ADF) Test: When the time series has a trend in it and is potentially slow-turning around a trend line you would draw through the data, the augmented dickey-fuller test equation by [Dickey and Fuller \(1981\)](#):

$$\nabla x_t = \theta_0 + \beta t + (\phi - 1)x_{t+1} + \theta_1 \nabla x_{t+1} + \dots + \theta_p \nabla x_{t+p} + w_t \quad (2)$$

where: ∇x_t is the first differenced value of series (x_t); w_t is the error term; x_{t-1} is the first lagged value of the series (x_t); ∇x_{t-j} is the j th lagged of the first differenced of values of x_t , $\theta_0, \beta, \phi, \theta_1, \theta_2, \dots, \theta_p$ are parameters to be estimated.

2.3.1.2. Building ARIMA model. As proposed by [Box and Jenkins \(1976\)](#) methodology, in order to build ARIMA model for a particular time series data, should follow four phases: Model identification, Estimation of model parameters, Diagnostic checking for the identified model, Application of the model (forecasting).

Model Identification: With this stage, the number of differencing required achieving stationary and the order of both the seasonal and non-seasonal AR and MA operators is determined. The autocorrelations function (ACF) and the partial autocorrelation functions (PACF) are the two most useful tools in any attempt at time series model identification ([Granger & Newbold, 1986](#)). To determine the number of differencing (d), non-seasonal autoregressive (p) and moving average (q) parameters, the guideline stated in ([Shumway & Stoffer, 2010](#)) are used. Accordingly, if PACF Cuts off after lags q and ACF tail off, then ARIMA (0, d, q) model is identified. If ACF cutoff after lag p and PACF tail off, ARIMA (p,d, 0) model is obtained. Finally, if both ACF and PACF tail off, then the identified model will be ARIMA (p,d,q).

Parameter Estimation: After choosing the most appropriate ARIMA model, the parameters are estimated by using Maximum Likelihood Estimation.

Diagnostic Checking: This stage, it deals with the residual assumptions in order to determine whether the residuals from fitted model are independent, constant variance, and normally distributed ([Shumway & Stoffer, 2010](#)).

- **Testing for independence:** Following technique which help us to checking independence of residual are applied:
 - > **Visual Analysis:** It is an approach for analyzing plot of the residual over time. If visual inspections of the plot reveal that they are randomly distributed over time, then it is a residual independence ([Shumway & Stoffer, 2010](#)).
 - > **Residual Autocorrelation Function (RACF):** With this test, to say that residual follows a white noise process, roughly 95% of the autocorrelation coefficient should fall within the range of $\pm 1.96/\sqrt{n}$ ([Lehmann & Rode, 2001](#)).
- **Test for normality:** The histogram and Q-Q plot used for visual analysis related to residual normality. If the histogram is bell shaped, that is most of the data values are at the center and the most of the value at QQ-plots lies at the reference line, then the data is normal.
- **Test of homoscedasticity:** Checking of residuals homoscedasticity, the Goldfeld-Quandt tests were used.

If the constant variance and normality assumptions are not true, they are often reasonably well satisfied when the observations are transformed by a Box-Cox transformation. The models that do not fulfill at least one of the diagnostic checks will be eliminated. And if the selected model is inadequate, the three-step model building process with other model is typically repeated several times until a satisfactory model obtained. The final selected model can then be used for prediction purposes ([Wei, 2006](#); [Shumway & Stoffer, 2010](#)).

Forecasting: It is last step in time series modeling, the goal is to predict future values of a time series, x_{t+m} , $m = 1, 2, \dots$ based on the data collected to the present, $x = \{x_t, x_{t-1}, \dots, x_1\}$.

$$x_{t+m} = \phi_1 x_{t+m-1} + \dots + \phi_{p+d} x_{t+m-p-d} - \theta_1 w_{t+m-1} - \dots - \theta_q w_{t+m-q} \quad (3)$$

3. Results and discussions

3.1. Descriptive analysis results

As seen from [Fig. 1](#), Ethiopia reported its first confirmed COVID-19 case on 14 March 2020 with one Corona positive person came from Japan. Then after, although the government of Ethiopia has immediately taken different measures to prevent and control the pandemic, the case has been rapidly increasing and widely distributing. Despite the fact that it seems slowly growing as compared to other East African countries, the country's daily record increase on average from 100 to 200 cases with the exception of highest cases, 399, which was recorded on 21 June 2020, to June 30, 2020. During the time of study, the number was too much enlarged, and more than six thousand (5,846) people were infected by coronavirus in Ethiopia. Moreover, as observed in [Table 1](#), the rate of new infection of COVID-19 in Ethiopia per 1000 high risky population was increased over time from 6.39 in March 2020 to 10.62 at May, 202, which indicated the potential but slow growing of the pandemic with time. However, higher rate of new infection (47.86) was observed by the month of June 2020. Likewise, the analysis showed that at the end of June, after four months since the first COVID-19 cases were reported in Ethiopia, the population level prevalence of COVID-19 was increased to 5.246% from 0.02% in March 2020. This indicate that relatively slow rate of infection of the disease over the population for the last four months.

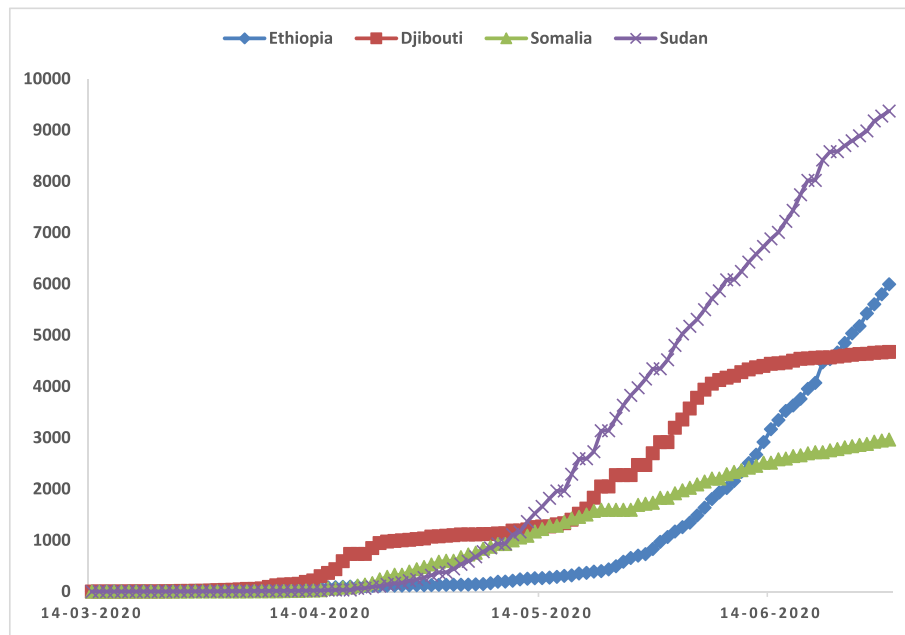


Fig. 1. Distribution of cumulative COVID-19 case in Ethiopia, Sudan, Djibouti, Somalia (13 Mar–June 30, 2020).

Table 1

Cumulative new cases, new infection & prevalence rate per 1000 population for COVID-19 cases, March to June 30, 2020.

Month	Ethiopia			Djibouti			Somalia			Sudan		
	CNC	NIR	PR (%)	CNC	NIR	PR (%)	CNC	NIR	PR (%)	CNC	NIR	PR (%)
March	23	6.39	0.02	31	8.56	3.14	3	3.99	0.019	7	3.5	0.016
April	130	8.92	0.114	1089	10.8	110.22	598	8.79	3.78	368	7.84	0.855
May	1063	10.62	0.933	3194	37.8	323.28	1318	10.62	12.06	4425	82.13	10.95
June	5812	47.86	5.246	4672	86.2	472.87	1647	36.2	18.66	4571	96.86	21.37

Key: CNC = cumulative new cases; NIR = new infection (incident) rate; PR(%) = percentage of monthly prevalence rate per respective population.

Despite for population size of the studied countries, the rate of infection was very slow in Ethiopia even as compared to Somalia, Djibouti, and Sudan from March through end of June of the study period. Moreover, very high increasing rate of infection prevalence was observed in Djibouti's (472.87% at the end June from 3.14% of March 30, 2020) than other countries for the past four months. This Result may be the very small size of population (1 million) of Djibouti than other.

Sudan, Djibouti and Somalia reported the first confirmed COVID-19 case on 15, 19, 17 Feb 2020. Each of the countries reported one confirmed at the time. From Fig. 1, it was observed that the trend of COVID-19 infection in Sudan, Djibouti and Somalia has continuously been increased, and the highest number people were infected by coronavirus despite the rate of infection was various. Until the end of June, Sudan has confirmed (9271), Djibouti has confirmed (4672), and Somalia has confirmed (2965), while the preventing and controlling mechanisms have been under practice.

3.2. Result from ARIMA modelling of COVID-19 case

Result in testing stationary: The first step in any time series analysis including ARIMA modeling is to check whether the time series is stationary. The stationary of a time series is important as easier to get accurate estimates (Elevli et al., 2016). In this study, time series plot, Autocorrelation Function (ACF) and Augmented Dickey-Fuller (ADF) test were applied to check the stationary. Accordingly, the pattern of the series of COVID-19 confirmed cumulative cases of East Africa Countries (Ethiopia, Djibouti, Somalia and Sudan) in Fig. 1 shows that there have been an obvious systematic upward change over the mean.

According to Shumway & Stoffer, 2010, if sample autocorrelation functions plot decay slowly in non-seasonal lags and took long lag to fall inside 95% significant line, it is an indication for non-seasonally non-stationary of the series. As it was observed in Fig. 2, sample autocorrelation function is also supporting evidence for the presence of regular linear trend behavior of all considered series so as it shows decaying slowly in non-seasonal lags to fall down inside 95% significant line.

The study also further tested the null hypothesis, which has a unit root with Augmented Dickey-Fuller (ADF) test using the test equation expressed in Eqn. (2) above. The test was used in order to investigate COVID-19 cases of Ethiopia, Djibouti, Somalia and Sudan whether stationary or not, at level (without differencing), and after first and second regular differencing,

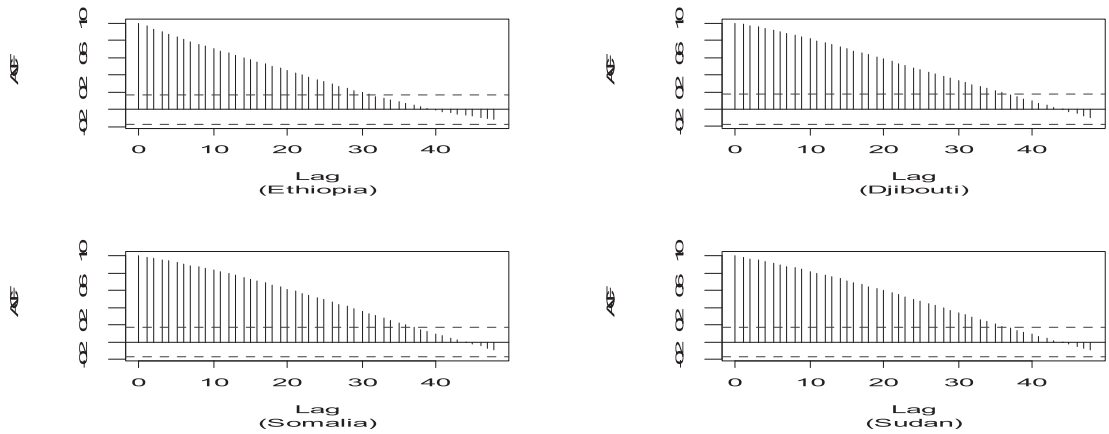


Fig. 2. Autocorrelations Function (ACF) plot of COVID-19 cases for selected East African countries.

and results were summarized in Table 2. According to these results, the null hypothesis, which has a unit root, for the data sequences should not be rejected at level and first differencing for all series except Djibouti (Stationary at first differencing). This figure further confirmed that original series (at level) as well as series obtained after first regular differencing for all series were in favor of visual analysis (time series plot and autocorrelation plot). Consequently, from both visual analysis, and ADF test, it can be decided that all data need second order non-seasonal differencing so as to make stationary in mean. However, according to Goldfeld-Quandt test (test-value = 12.35, P-value = 0.03, 0.01, 0.00, 0.034), coronavirus case series of selected East Africa (Ethiopia, Djibouti, Somalia and Sudan, respectively) was not acknowledged as stationary in variance during study period. Hence, it required to take remedy approaches to stabilize series stationary in variance. And, it applied natural log transformation all series and consequently, variance were stabilized. Thus, COVID19 case series of Ethiopia, Djibouti, Somalia and Sudan became stationary in mean and stabilized in variance after the second-order differencing and log transformation (see Fig. 4 below and Appendix Fig. A1). From this, a candidate ARIMA models were recommended for all considered series based on autocorrelation function (ACF) and partial autocorrelation (PACF) plot of Fig. 3. See the results in Table 2 below;

According to Shumway & Stoffer, 2010, all the models that passed and fulfilled all residual tests (independence, normality and homogeneity) and the parameters significantly differ from zero must be included and selected as candidate model for

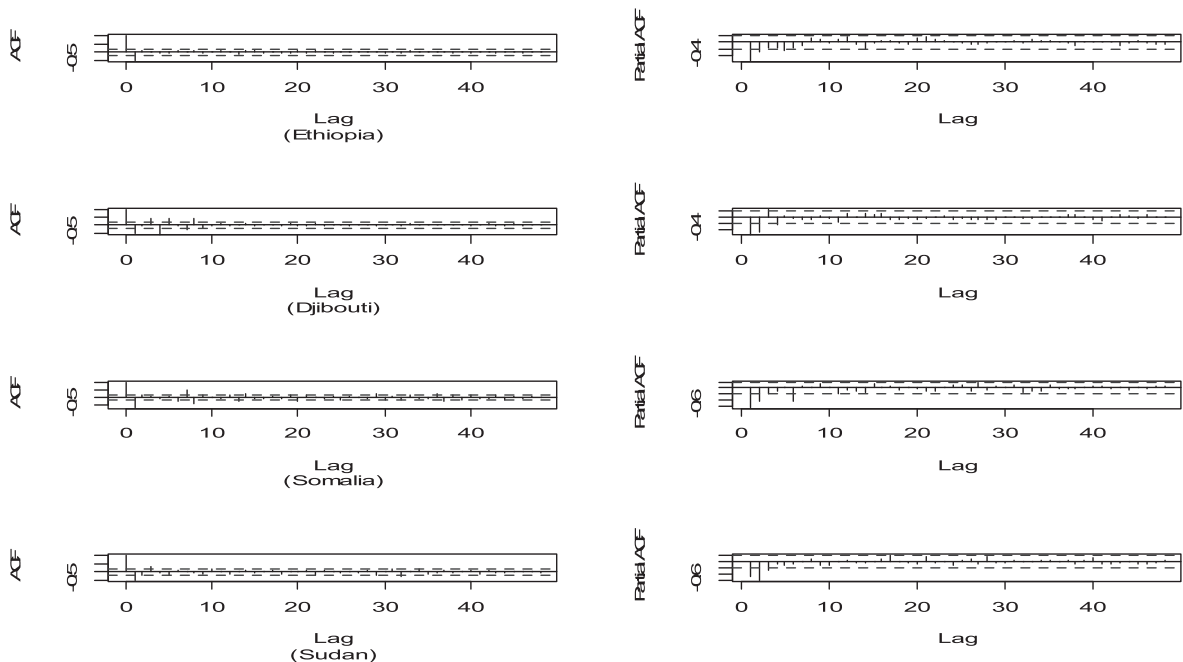


Fig. 3. Estimated ACF and PACF plot to predict the trend of COVID-19 cases in selected East Africa countries after second-order differencing.

Table 2

Summary of ADF unit-roots test (in level and after first and second-order regular differencing).

Series level	Ethiopia	Djibouti	Somalia	Sudan	Decision: Null at unit root (non-stationary)
	ADF test statistic (<i>p-value</i>)				
Original series	-0.781 (0.12)	-1.442 (0.27)	-1.976 (0.12)	-2.123 (0.52)	Don't Reject null hypothesis
First regular differencing	-0.412 (0.07)	-3.44 (0.00)	-3.56 (0.07)	-5.682 (0.13)	Don't Reject null but Accepted Djibouti series
Second-order differencing	-4.97 (0.01)		-2.87 (0.002)	-3.121 (0.01)	Reject null hypothesis

prediction. Then, all candidate models are compared their AIC and BIC, and the one which has least will be selected as the best model for prediction model.

Accordingly, among the candidate models, ARIMA (1,2,1), ARIMA (2,1,1) with drift, ARIMA (0,2,2) and ARIMA (2,2,1) models were chosen as the best models for predicting coronavirus cases of Ethiopia, Djibouti, Somalia, and Sudan respectively. This was because it fulfills all diagnostic tests (independence, normality and homogeneity) and has least AIC and BIC (See Table 3, and Appendix Fig. A3 and A4).

Forecasting accuracy evaluation

If the fitted models perform well in forecasting, the forecast error will be relatively small and the Mean Absolute Percentage Error (MAPE) should be close to 4%. From Table 3, it can be observed that the accuracy of forecasts measured by the Mean Absolute Percentage Error (MAPE) turned out to be 3.92% for ARIMA (1, 2, 1) model, 3.71% for ARIMA (2, 1, 1) with drift model, 3.59% for ARIMA (0, 2, 2) model and 3.92% for ARIMA (2, 2, 1) model, which are relatively less than 4%. This implies that those model would perform well in predicting the coronavirus cases of the selected East Africa Countries.

The actual and predicted values with 95% prediction interval using best prediction models of COVID-19 cases for all studied East Africa countries were investigated, and Table 4 presents the scenario forecasts for the following four months in Ethiopia, Djibouti, Somalia and Sudan (from July 1 until October 30, 2020) using the results of Box-Jenkins ARIMA model. The results were categorized into two future scenarios - worst-case scenario, and optimistic/average-case scenario. It is assumed that the worst-case scenario happens if and only if the standard WHO disease preventive and controls measures are not taken. Contrarily, the preventive and control measures are practiced seriously when optimistic-case scenario happens. Table 4 and Fig. 4 presents cumulative COVID-19 cases and prevalence rates of East Africa countries where the number of cumulative COVID-19 infected cases is expected to considerably increase in the coming four months in all study areas. By the end of October, cumulative infections across Ethiopia, Djibouti, Somalia and Sudan will reach 56,610, 8,336, 3,961, and 21,388 respectively when looked at an average case scenarios.

Specifically in Table 4: Panel A, the estimates suggest that at the end of July, cumulative COVID-19 infections in Ethiopia will reach 15,169 (CI 95% 14,112 to 16,226), and 42,797 (CI 95% 26,314 to 59,279) on September 30, 2020. And, this will further increase to 56,610 (CI 95% 28,723 to 84,497) by October 30, 2020. This means that in the coming four months, the number of coronavirus positive people in Ethiopia may lift up to 56,610 from the current number, 5,846 (on June 30, 2020), on average-case scenario. However, in worst case scenario forecast by this model, the number of infections will be around 84,497 after 4 months. This further suggests, the infected cases will be increased by 3–4 times than the current figures for the next four month if effective preventive measures will not be taken seriously (such as quarantine all cases in an attempt to cut off the source of infection, close contact tracing, population movement control especially border area, government and people intervention. Moreover, medically overstretched phenomenon becoming serious in Ethiopia, and there will emerge the large number of death toll.

Table 3

Candidate ARIMA models & coefficients for best selected model.

Country	Model	AIC	BIC	MAPE	Coefficient	P-value
Ethiopia	ARIMA (0,2,2)	1090	1098			
	ARIMA (1,2,0)	1070	1082			
	ARIMA (1,2,1)	1041.14	1051.68	3.92%	$\hat{\phi}_1 = 0.8651, \hat{\theta}_1 = -1.8746, \hat{\theta}_2 = 0.9391$	0.000,0.00,0.01
Djibouti	ARIMA (2,1,1) with drift	1048.84	1059.18	3.71%	$\hat{\phi}_1 = -1.2315, \hat{\phi}_2 = -0.7172, \hat{\theta}_1 = 0.5868$	0.000,0.000,0.00
	ARIMA (0,2,1)	1094.12	1098.2			
	ARIMA (5,2,2)	1128.47	1158.23			
Somalia	ARIMA (0,2,2)	919.44	927.26	3.59%	$\hat{\theta}_1 = -1.0169, \hat{\theta}_2 = 0.2577$	0.001,0.000
	ARIMA (2,2,0)	1442	1124			
	ARIMA (0,2,1)	1234	1112			
Sudan	ARIMA (1,2,2)	1285.3	1201			
	ARIMA (2,2,1)	1162.22	1172.72	3.92%	$\hat{\phi}_1 = -0.5071, \hat{\phi}_2 = -0.4555, \hat{\theta}_1 = -0.5290$	0.000,0.01,0.00
	ARIMA (0,2,1)	1310	1285			

Table 4
Prediction of COVID-19 infected cases for 4 months Ahead, July until October 30, 2020.

Panel-A: Cumulative case in by the end of each month: Forecast, July–Oct 30, 2020												
Country	Point forecast (by the end of each month)				Lower prediction Limit (95% PI)				Upper prediction limit (95% PI)			
	July	Aug	Sep	Oct	July	Aug	Sep	Oct	July	Aug	Sep	Oct
Ethiopia	15,169	28,983	42,797	56,610	14,112	21,701	26,314	28,723	16,226	36,265	59,279	84,497
Djibouti	5,216	6,196	7,264	8,336	4,744	4,399	4,433	4,585	5,938	8,944	11,592	14,072
Somalia	3,212	3,462	3,711	3,961	3,041	2,399	1,346	0	3,475	5,088	7,328	10,037
Sudan	12,044	15,159	18,274	21,388	11,530	12,172	11,708	10,413	12,831	19,727	28,315	38,174

Panel-B: Cumulative infection rate per 1,000 population in each month: Forecast, July–Oct 30, 2020												
Country	Point forecast (by the end of each month)				Lower prediction Limit (95% PI)				Upper prediction limit (95% PI)			
	July	Aug	Sep	Oct	July	Aug	Sep	Oct	July	Aug	Sep	Oct
Ethiopia	0.13	0.26	0.38	0.49	0.12	0.19	0.23	0.25	0.14	0.32	0.52	0.74
Djibouti	5.3	6.3	7.4	8.50	4.8	4.45	4.49	4.64	6.0	9.05	11.7	14.24
Somalia	0.2	0.22	0.23	0.25	0.19	0.15	0.09	0.00	0.22	0.32	0.46	0.63
Sudan	0.28	0.35	0.42	0.49	0.26	0.28	0.27	0.24	0.29	0.45	0.67	0.87

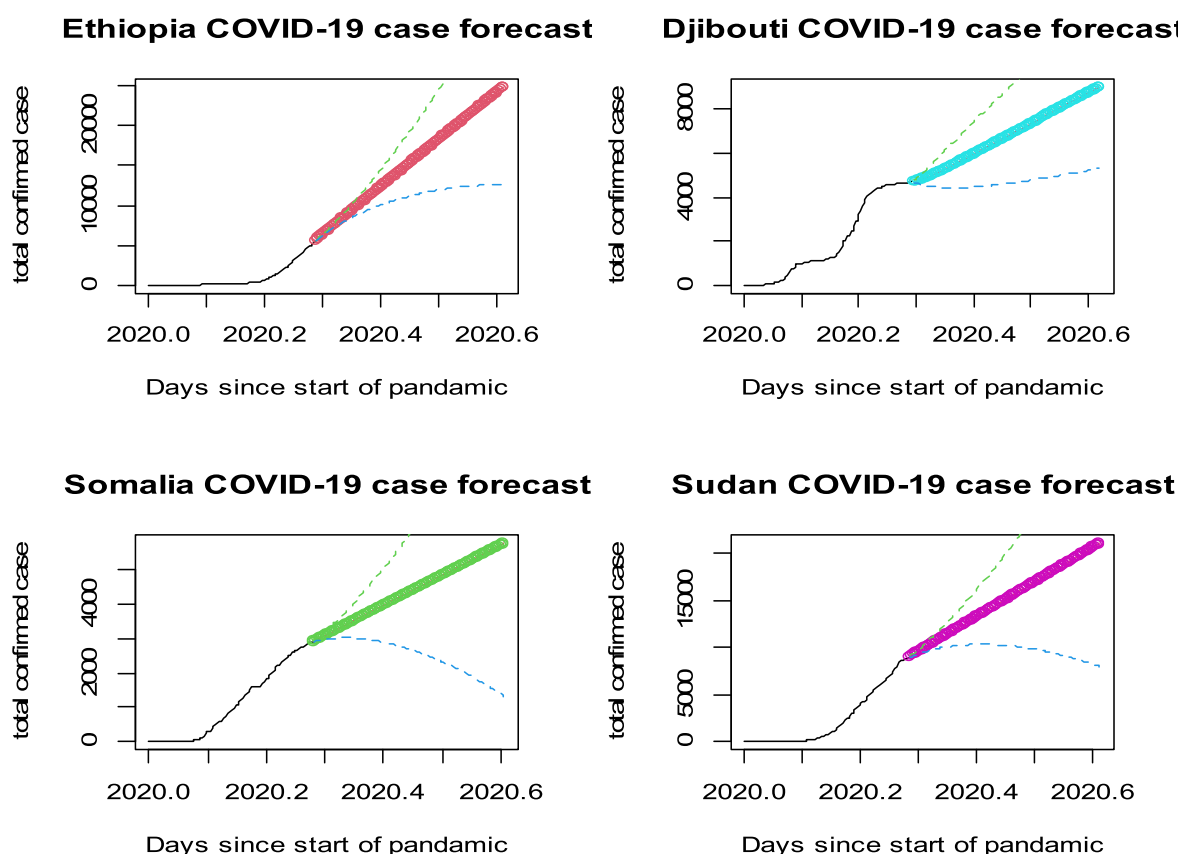


Fig. 4. Actual and forecasted value with 95% prediction Interval studied East African countries.

Likewise, looking over the coming 4-months’ scenario, the model depicted that with proper interventions and control, cumulative number of infected people in Djibouti, Somalia and Sudan will increase from 4,656, 2,904 and 9,258 at the end of June to 8,336, 3,961 and 21,388 respectively by October 30, 2020. If necessary preventive measures are not taken seriously right from this day, may really happens because this projections is alarming, and it is based on basic fundamentals of the Time Series Forecasting techniques. The number of corona positive people may increase very quickly and may reach to 14,072, 10,037 and 38,174 in Djibouti, Somalia and Sudan respectively by the end of October at the worst case scenario.

The study further estimated cumulative infection (prevalence) rates as shown in Panel B of Table 4 by combining the forecasts on cumulative cases with UN population estimates. The Result suggested that population level prevalence rate of

COVID-19 in Ethiopia is expected to increase from 5.3% by the end of June 2020 to 49% in October 30, 2020 (CI 95% 25%–74%) throughout the prediction period. The cumulative infection rate of Djibouti, Somalia and Sudan after 4 months will be 850%, 25% and 49% respectively than 472.9%, 18.7%, 21.4% by June 30, 2020.

3.3. Discussions of present finding

It is essential to create a reliable and suitable predictive model that can help governments and other stakeholders to control the further spread of Novel coronavirus. Time series forecasting models is the statistical technique which gives a decent predictions and has been widely applied for trend of infectious disease in quick time (Dehesh et al., 2020; Grasselli, Pesenti, & Cecconi, 2020; Liu, Beeler, & Chakrabarty, 2020; Liu, Magal, Seydi, & Webb, 2020; Shi & Fang, 2020; Wongkoon et al., 2012; Wu et al., 2020). In this study, with the view to predicting coronavirus (COVID-19) prevalence in East African, mainly, Ethiopia, Djibouti, Somalia and Sudan, using ARIMA modeling strategy, which would be a useful guidance for timely prevention and control measure to be effectively planned in advance.

The current situations of coronavirus in Ethiopia, Djibouti, Somalia and Sudan, with assume that all people are at risk of this pandemic, has determined and presented incidence rate and prevalence rate, as seen in Table 1. Through strict follows of all stage Box-Jenkins strategy, ARIMA (1,2,1), ARIMA (3,1,1) with drift, ARIMA (0,2,2) and ARIMA (2,2,1) models were chosen as the optimal models for predicting coronavirus cases of Ethiopia, Djibouti, Somalia, and Sudan respectively. By means of those model, a forecasts of four month a heads future scenario COVID-19 prevalence (July until October 30, 2020) has made. And it was outlooks with two scenario categories as the worst-case and optimistic/average-case scenario. According to the results, the incidence of the COVID-19, shows an increasing growth steep in all considered area during study period. The spreading trend of the virus is expected to move in upward trend for coming four month a head from July 2020, which peak cannot said to be reached yet.

In different area, a lot of empirical studies have been conducted to predict prevalence of COVID-19 case using ARIMA model (Fanelli & Piazza, 2020; Gupta & Pal., 2020; Liet al., 2020; Roosa et al., 2020; Yonar et al., 2020). Consistent with this study, Zeynep, 2020 found that ARIMA models appropriate for predicting the prevalence of COVID-19 infectious disease and they selected ARIMA (0,2,1), ARIMA (1,2,0), and ARIMA (0,2,1) models for Italy, Spain, and France, respectively. By means of those model, they estimated the ongoing trend and extent of the outbreak of coronavirus. In the other hand, Gupta & Pal., 2020 using ARIMA model analyzing and Forecasting of COVID-19 outbreak in India and shows that the infected cases in India in worst case scenario, average and most optimistic scenario; in the way that will used for the key stakeholders like India Government so as to prepare a combat plan with rigorous measures. Even if different model is obtained from our study, they show that ARIMA model effectively able to handle the prevalence of the new coronavirus.

4. Conclusion

The trajectory of the cumulated infection of COVID-19 pandemic is forecasted by using ARIMA model strategy so as to help governments and other stakeholders to control the further spread of Novel coronavirus. The Result of this study is an alert of the possibilities of COVID-19 spreading more than it's currently observed from the model prediction. If the incidence rate of the virus not change very strangely, these models prediction are optimal and could be helpful and indicate the future scenarios of the disease. However, it already recognized that COVID-19 is new pandemic disease and have the ability to be transmitted rigorously and may affect all the prediction.

Therefore, an effective implementation of the preventive measures and a rigorous compliance with the rules such as prohibiting public gatherings, travel restrictions, personal protection measures, and social distancing may alleviate the spreading rates of the virus, particularly, Sudan and Ethiopia. Moreover, more efforts should be exerted on Ethiopian side to control the population movement across all the border areas and to strengthen border quarantining.

Further, continuous reconsideration of predictive model by updating more new data, provide useful and more precise prediction. Apply, ARIMAX-Transfer Function model in region-wise by take in to consideration of climatic factor like temperature and humidity in each countries regarding the spatial pattern for reliable measure of COVID-19 prevalence, since recently the virus distributed at the community level.

Declaration of competing interest

I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the John Hopkins University for publicly releasing the updated datasets on the number of infected cases of COVID-19.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.idm.2020.08.005>.

References

- Abivenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). *Application of the ARIMA model on the COVID-2019 epidemic dataset*, Article 105340. Data in brief.
- Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modeling and forecasting of the COVID-19 outbreak. *PLoS One*, *15*, Article e0230405. <https://doi.org/10.1371/journal.pone.0230405>
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden day.
- Dehesh, T., Mardani-Fard, H. A., & Dehesh, P. (2020). *Forecasting of COVID-19 confirmed cases in different countries with ARIMA models*. medRxiv.
- Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *1981*, 1057–1072.
- Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, *134*, 1–12. <https://doi.org/10.1016/j.chaos.2020.109761>
- Granger, K., & Newbold, J. (1986). *Forecasting economic time series*. USA: Academic Press.
- Grasselli, G., Pesenti, A., & Cecconi, M. (2020). Critical care utilization for the COVID-19 outbreak in lombardy, Italy: Early experience and forecast during an emergency response. *Journal of the American Medical Association*, *323*(16), 1545–1546.
- Gupta, R., & K Pal, S. (2020). Trend analysis and forecasting of COVID-19 outbreak in India. *MedRxiv*. <https://doi.org/10.1101/2020.03.26.20044511>. preprint.
- Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). *Artificial intelligence forecasting of covid-19 in China*. arXiv preprint arXiv 2002.07112.
- Jia, W., Han, K., Song, Y., Cao, W., Wang, S., Yang, S., & Liu, M. (2020). Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *medRxiv*.
- John Hopkins University. (2020). Novel coronavirus (COVID-19) cases, provided by JHU CSSE Accessed from <https://github.com/CSSEGISandData/COVID-19.on.june.2020>.
- Kim, S. K. (2020). AAEDM: theoretical dynamic epidemic DiffusionModel and covid-19 Korea pandemic cases. *medRxiv*.
- Lehmann, A., & Rode, M. (2001). Long-term behaviour and cross-CorrelationWater quality analysis of the river elbe, Germany. *Journal of Water Resources*, *35*, 2153–2160.
- Li, Q., Feng, W., & Quan, Y. H. (2020). Trend and forecasting of the COVID-19 outbreak in China. *Journal of Information Security*, *80*, 469–496. <https://doi.org/10.1016/j.jinf.2020.02.014>
- Liu, P., Beeler, P., & Chakrabarty, R. K. (2020a). COVID-19 progression timeline and effectiveness of response-to-spread interventions across the United States. *medRxiv*.
- Liu, Q., Liu, X., Jiang, B., & Yang, W. (2011). Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infectious Diseases*, *11*(1), 218.
- Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020b). *Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data*. arXiv preprint arXiv 2002.12298.
- Lover, A. A., & McAndrew, T. (2020). Sentinel event surveillance to estimate total SARS-CoV-2 infections, United States. *medRxiv*.
- Luz, P. M., Mendes, B. V., Codeço, C. T., Struchiner, C. J., & Galvani, A. P. (2008). Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *The American Journal of Tropical Medicine and Hygiene*, *79*(6), 933–939.
- Massonnaud, C., Roux, J., & Crépey, P. (2020). COVID-19: forecasting short term hospital needs in France. *medRxiv*.
- Rios, M., Garcia, J. M., Sanchez, J. A., & Perez, D. (2000). A statistical analysis of the seasonality in pulmonary tuberculosis. *European Journal of Epidemiology*, *16*(5), 483–488.
- Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., Yan, P., & Chowell, G. (2020). *Infectious disease modelling*. 5 pp. 256–263). <https://doi.org/10.1016/j.idm.2020.02.002>
- Russo, L., Anastassopoulou, C., Tsakris, A., Bifulco, G. N., Campana, E. F., Toraldo, G., & Siettos, C. (2020). Tracing day-zero and forecasting the fade out of the COVID-19 outbreak in lombardy, Italy: A compartmental modeling and numerical optimization approach. *medRxiv*.
- Shi, Z., & Fang, Y. (2020). Temporal relationship between outbound traffic from Wuhan and the 2019 coronavirus disease (COVID-19) incidence in China. *medRxiv*.
- Shumway, R. H., & Stoffer, D. S. (2010). *Time series analysis and its applications with R Examples* (3rd ed.). Springer.
- Wei, W. S. W. (2006). *Time series analysis univariate and multivariate* (p. 478). New York-USA: Addison-Wesley Publishing Company, Inc..
- Wise, T., Zbozinek, T. D., Michelini, G., & Hagan, C. C. (2020). *Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States*.
- Wongkoon, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2012). Development of temporal modeling for prediction of dengue infection in Northeastern Thailand. *Asian Pacific journal of tropical medicine*, *5*(3), 249–252.
- World Health Organization. (2020). Coronavirus disease (COVID-19) pandemic, WHO Accessed from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019.on.june.2020>.
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet*, *395*, 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
- Yonar, H., Yonar, A., Tekindal, M. A., & Tekindal, M. (2020). Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the box-jenkins and exponential smoothing methods. *EJMO*, *4*(2), 160–165.
- Zeynep, C. (2020). *Estimation of COVID-19 prevalence in Italy, Spain, and France samsun university*. Turkey: Faculty of Engineering, Industrial Engineering Department, 55420S amsun. <https://doi.org/10.1016/j.scitotenv.2020.13881>
- Zhan, C., Chi, K. T., Lai, Z., Hao, T., & Su, J. (2020). Prediction of COVID-19 spreading profiles in South Korea, Italy and Iran by data-driven coding. *medRxiv*.