

RESEARCH

Open Access



Short-term forecasting of daily infections, fatalities and recoveries about COVID-19 in Algeria using statistical models

Firdos Khan¹ and Mohamed Lounis^{2*}

Abstract

Background: A viral disease due to a virus called SARS-Cov-2 spreads globally with a total of 34,627,141 infected people and 1,029,815 deaths. Algeria is an African country where 51,690, 1,741 and 36,282 are currently reported as infected, dead and recovered. A multivariate time series model has been used to model these variables and forecast their future scenarios for the next 20 days.

Results: The results show that there will be a minimum of 63 and a maximum of 147 new infections in the next 20 days with their corresponding 95% confidence intervals of – 89 to 214 and 108–186, respectively. Deaths' forecast shows that there will be 8 and 12 minimum and maximum numbers of deaths in the upcoming 20 days with their 95% confidence intervals of 1–17 and 4–20, respectively. Minimum and maximum numbers of recovered cases will be 40 and 142 with their corresponding 95% confidence intervals of – 106 to 185 and 44–239, respectively. The total number of infections, fatalities and recoveries in the next 20 days will be 1850, 186 and 1680, respectively.

Conclusion: The results of this study suggest that the new infections are higher in number than recover cases, and therefore, the number of infected people may increase in future. This study can provide valuable information for policy makers including health and education departments.

Keywords: Algeria, Time series model, Forecasting, COVID-19, Infections, Recoveries

1 Background

Since December 2019, the world is facing a new respiratory disease called coronavirus disease 2019 (COVID-19). This viral respiratory illness was first detected in the China's city of Wuhan in the province of Hubei [1]. On January 30, 2020, the World health organization (WHO) has declared it as an epidemic of "public health emergency of international concern." Forty days later, on March 11th, it has finally recognized this disease as a global pandemic urging all countries to intensify their efforts to prevent its propagation stressing the importance of

detection, testing, treatment, isolation, tracing and people mobilization in the anti-COVID-19 response [2]. As the number of affected persons is more than 34 million cases and 1,029,815 deaths through the world until now [3], this disease has attracted a global interest and enormous numbers of researches are in continuous struggle to understand its epidemiological and clinical characteristics. In addition to medical and experimental studies, researchers are using mathematical and statistical models to predict the number of affected persons, the peak and the ending time of the epidemic. The predicted scenarios can be used to assess the preventive measures which could be of great importance for decision-makers to adopt the best strategies in the anti-COVID-19 battle [4].

In this way, multiple models have been proposed for modeling COVID-19 pandemic in different countries including compartmental models, natural growth model

*Correspondence: lounisvet@gmail.com

² Department of Agro-Veterinary Science, Faculty of Natural and Life Sciences, University of Ziane Achour, Road of Moudjbara, BP 3117, 17000 Djelfa, Algeria

Full list of author information is available at the end of the article

and logistic growth models [5]. The SIR (susceptible–infectious–recovered) and SEIR (susceptible–exposed–infectious–recovered) historical compartmental models and their variants are the most used in forecasting human epidemic diseases [6]. They are also widely used in the case of COVID-19 [7]. However, these models have failed so far to give a good description of the empirical data [8]. Time series models (TSMs) are autoregressive moving average models which attempt to predict future events by means of aggregating recent data [8]. These models have shown many successful implementations in economics, finance, climatology, hydrology engineering and epidemiology [9]. The autoregressive integrated moving average (ARIMA), long short-term memory networks (LSTM) are the most commonly used for forecasting epidemic diseases [10]. These models have recently been widely used to forecast COVID-19 epidemic [11–16].

The aim of this study is to identify the data generating process of COVID-19 infections, fatalities and recoveries for Algeria and to forecast future scenarios about these variables. To our knowledge, there is no such study so far to forecast each variable of COVID-19 in Algeria; therefore, the findings of this research may help policy makers to reshape their policies according to the predicted scenario of COVID-19. Further, the results of this study can help various stakeholders including health and education department etc., in making their future's plans.

2 Data and study area

Daily data of new confirmed cases, deaths and recoveries about COVID-19 over Algeria from February 25 to October 1, 2020, were collected from World Health Organization (WHO). In Algeria, the first case of COVID-19 was reported on February 25, 2020. This methodology is applicable to any other region of the world as long as you have time series data.

3 Methods

Time series models have multitude of applications in different areas from finance to climatology and hydrology and have shown good performances in forecasting future scenarios [17–19]. There are different classes of time series models, and they can be used to model time series data which depends on their nature. Due to the dependence nature of the considered three variables, it was decided to model these variables jointly. Consequently, VAR model was used for modeling and forecasting COVID-19 in Algeria. Suppose we have a vector of time series denoted by Y_t , then VAR model can be expressed by Eq. (1)

$$Y_t = B_0 + B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_p Y_{t-p} + \epsilon_t \quad (1)$$

where in Eq. (1), B_0 and ϵ_t are $k \times 1$ column vectors and B_i (for $i=1, 2, \dots, p$) is $k \times k$ matrices of coefficients. Y_t is a

vector of $k \times 1$ time series variables. To incorporate VAR model, a four-step methodology have been implemented and given below:

- i. model selection (lag length)
- ii. estimation of unknown parameters of identified model
- iii. diagnostic check of estimated model
- iv. forecasting

The first step in modeling a time series data with VAR model is the selection of lags. In the literature, there are some criteria available for this purpose like Hannan–Quinn criterion (HQC) [20], Akaike's information criterion (AIC) [21], prediction error (FPE) [22] and Schwarz criterion (SC) [23]. The second step consists on the estimation of unknown parameters of identified model in the first step. Among various methods for estimation, ordinary least square (OLS) method of estimation has been used in this study. In the third step, the estimated model is diagnosed by using various statistical and graphical tests. The residuals of the estimated model can be visualized for normality, serial correlation and autoregressive conditional heteroscedasticity (ARCH) error. Statistical tests like normality and serial correlation test can be used for checking the assumption of the estimated model. The final step in time series modeling is to forecast the future's phenomena on the basis of fitted model to available data. Generally, n -step ahead forecast can be made. However, this depends on the nature and objective of the study. For analysis and visualization, we have used R software with packages *mFilter* [24] and *vars* [25].

4 Results

The daily new infections, fatalities and recoveries are displayed in Fig. 1, where it is obvious that there is an increasing trend in different time durations and a decreasing trend in other durations in Algeria. Table 1 presents the correlation structure among COVID-19 infections, fatalities and recoveries in Algeria. It can be seen that there is a positive relationship among these variables, and therefore, these variables were modeled jointly. VAR model was used to describe the data generating process of all these variables where the lag length was selected by using model selection criteria like AIC, SC, HQC and FPE. The best model was selected with 18 lags. The unknown parameters of identified model were estimated by using OLS method of estimation. The residuals of fitted model were analyzed and checked for normality and serial correlation graphically as well as numerically. It was found that the model has no serial correlation and the residuals follow approximate normal distribution. In addition, the residuals of the fitted

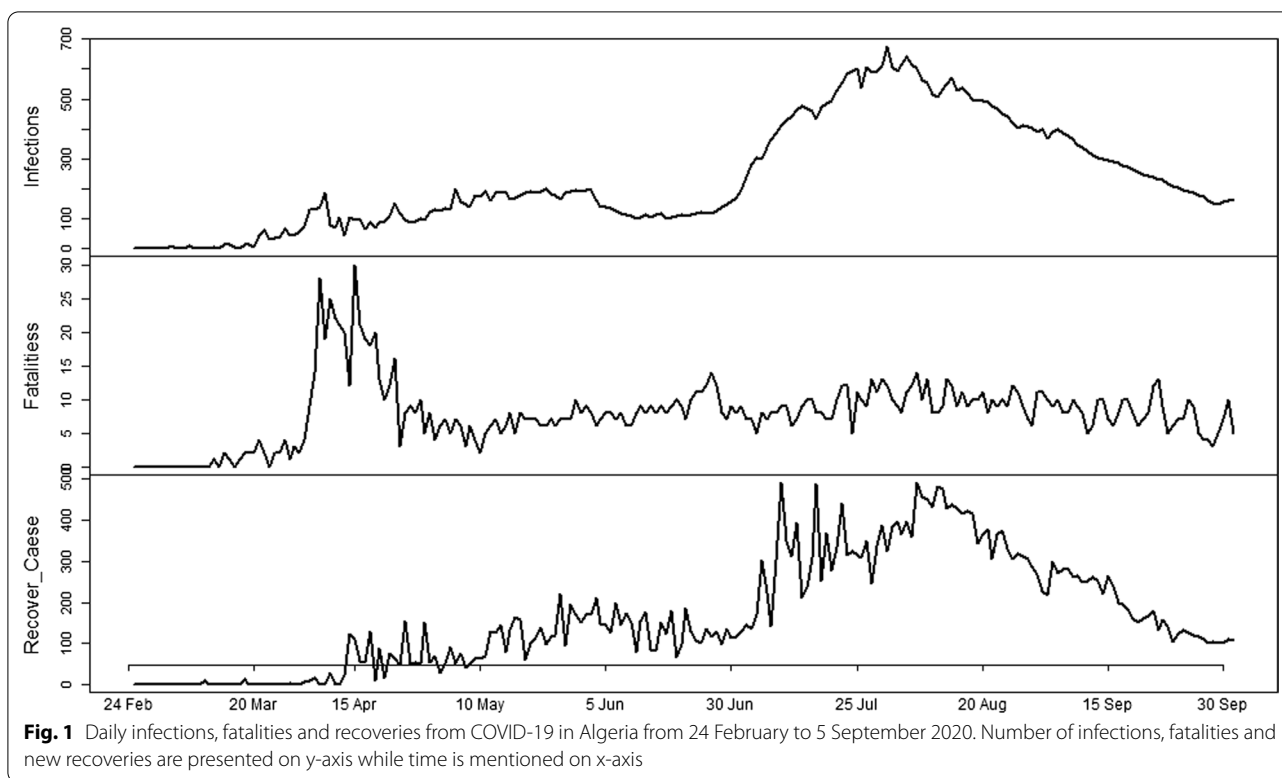


Table 1 Dependence structure among daily new cases, deaths and recoveries about COVID-19 in Algeria given by correlation matrix

	Infections	Fatalities	Recoveries
Infections	1.0000	0.3191	0.9206
Fatalities	0.3191	1.0000	0.2876
Recoveries	0.9206	0.2876	1.0000

model were investigated for ARCH error; however, the null hypothesis of ARCH error was rejected and it was concluded that there is no ARCH error. The fitted model was then used for 20 days ahead forecast of new cases, deaths and recovered cases of COVID-19 in Algeria.

The forecast results are presented in Fig. 2–4 and Table 2 where Figs. 2, 3 and 4 show results for new infections, fatalities and recover cases, respectively. Table 2 presents the forecast results of minimum, maximum and average values with their corresponding 95% confidence intervals for each variable of the forecast duration. The 20 days ahead forecast results for new cases of COVID-19 show that there is a decreasing trend. The maximum number of daily infections will be 147 with 95% confidence intervals of 108–186. The minimum number of daily infections forecast by our model is 63 with 95% confidence intervals

of – 89 to 214. Regarding fatalities, results for next 20 days show that their number will varied from 8 to 12 deaths per day, with a corresponding 95% confidence intervals of – 1 to 17 and 4–20, respectively. The model’s forecast results for daily recovered cases show that there will be a minimum of 40 [CI= – 106 to 185] and a maximum of 142 [CI= 44–239] recovered persons per day in the upcoming 20 days in Algeria. On the average, there will be 93, 10 and 84 new infections, fatalities and recoveries with their corresponding 95% confidence intervals of – 16 to 201, 2–18 and – 45 to 210, respectively.

The forecast results of VAR model for next 20 days show that the forecast values follow the trend of observed data reasonably. It can be seen from the forecast results that the average number of new recoveries is lower than the average number of new infections in next 20 days. Therefore, there will probably be more infected people in Algeria in the future. However, these results based on our model and the number of infected people may be reduced by following the SOPs of WHO and Algerian government about COVID-19.

5 Discussion

There are various studies about COVID-19 forecasting in different region and countries of the world [9, 15, 16, 19, 26–28] including Pakistan, Canada, Germany, UK, France, Italy, Russia, Japan and Turkey. The studies of Aslam

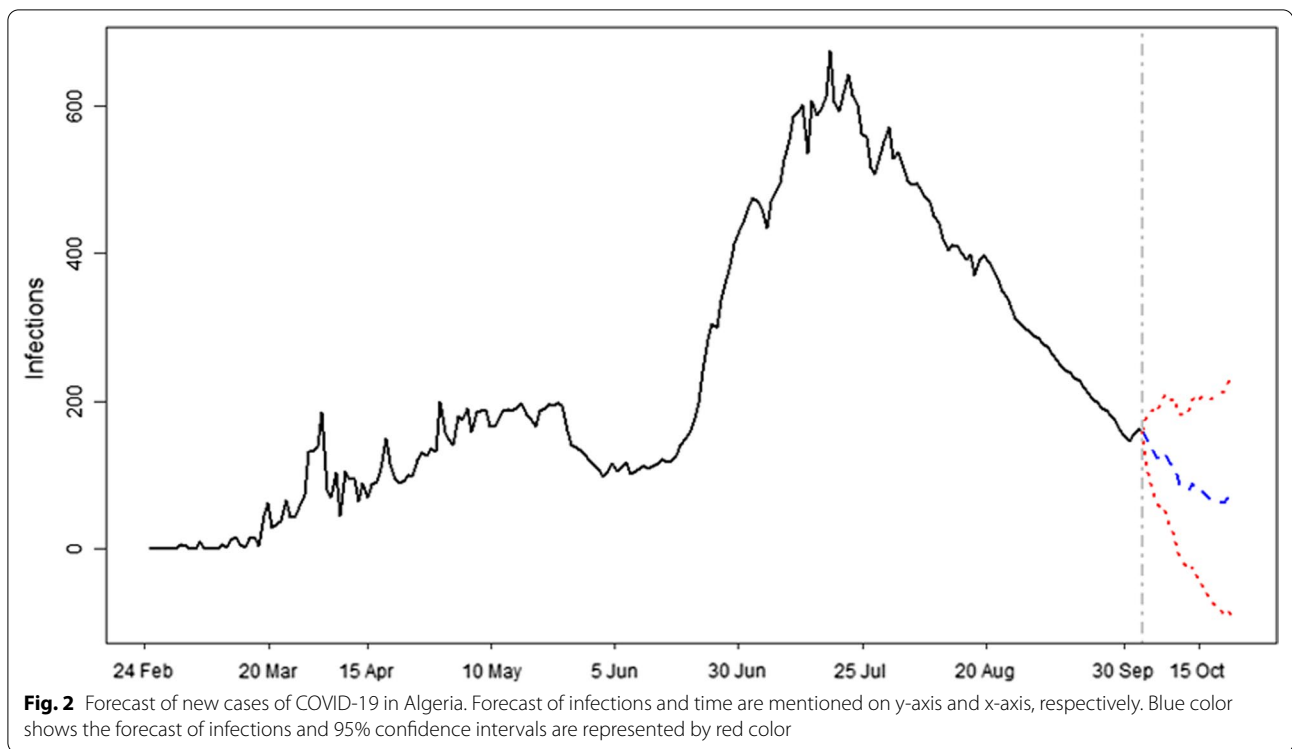
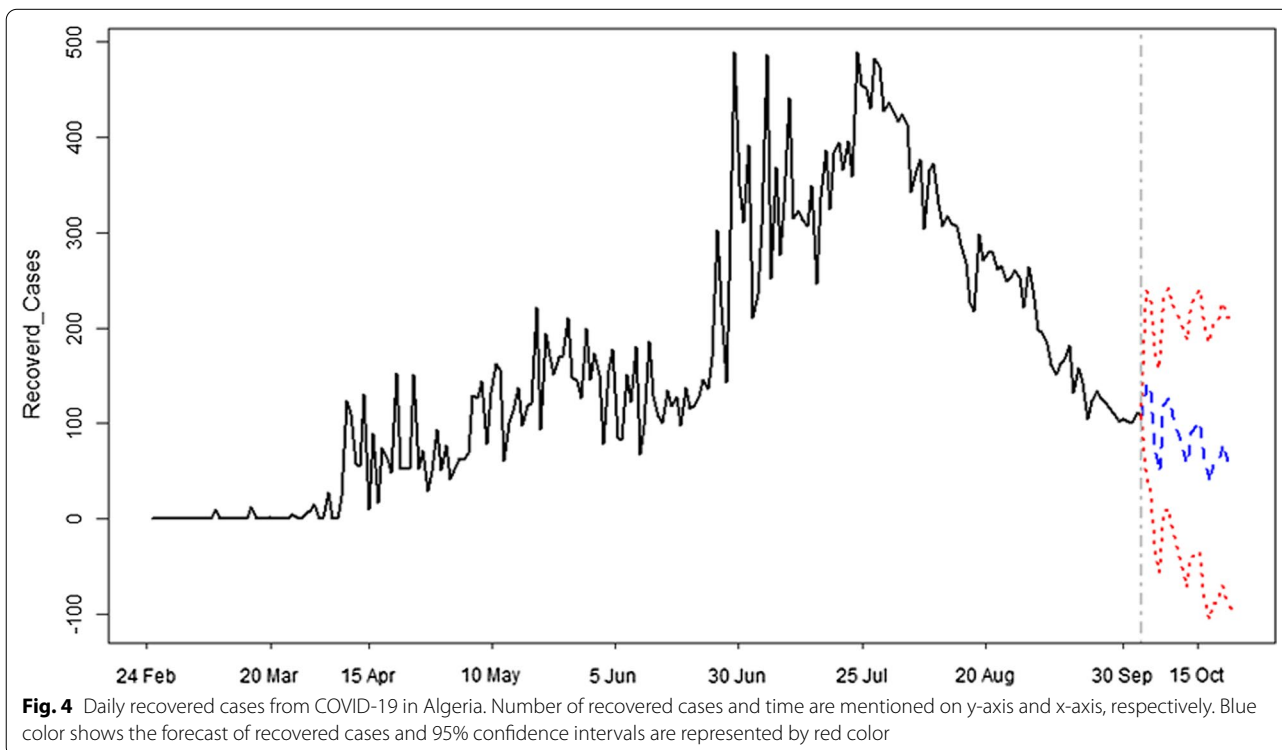
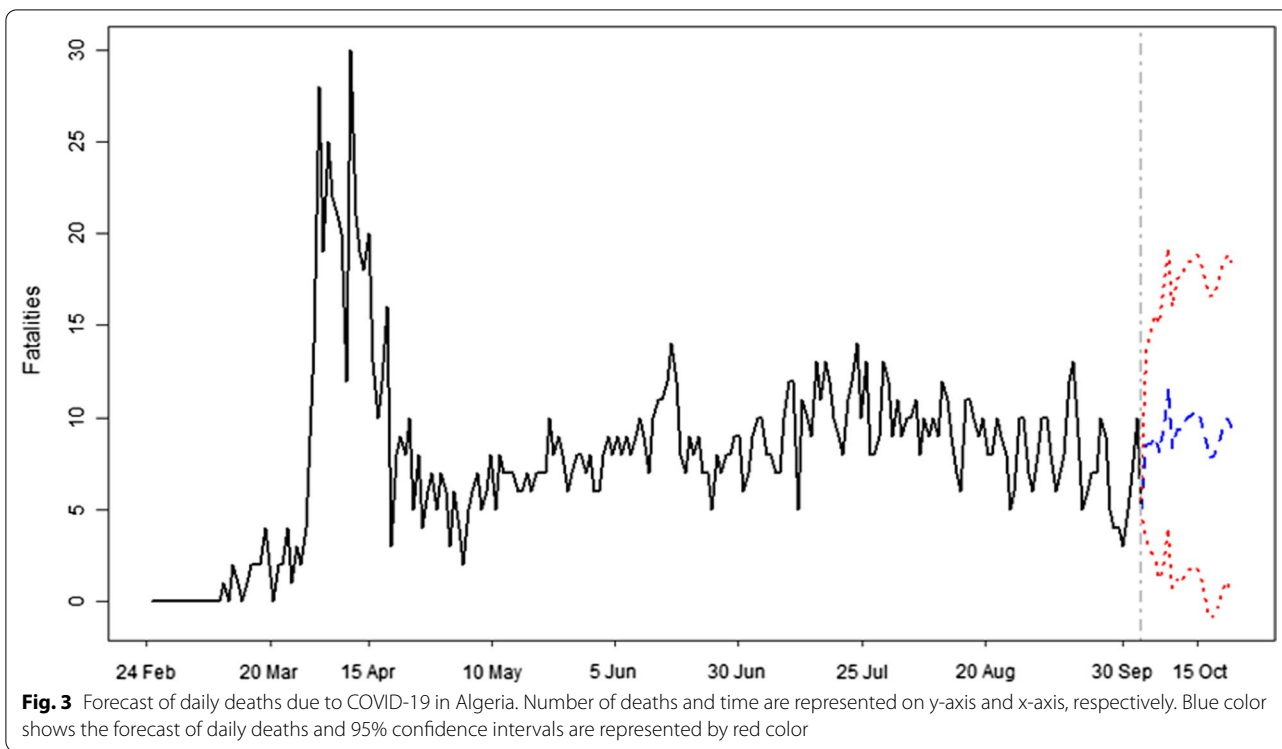


Table 2 Forecast values for daily infections, fatalities and recoveries from COVID-19 with their corresponding 95% confidence intervals for Algeria in upcoming 20 days

Variable	Max/Min	Forecast	95% confidence intervals		Total cases in 20 days
			Lower limit	Upper limit	
Infections	Min	63	-89	214	1850
	Max	147	108	186	
	Average	93	-16	201	
Fatalities	Min	8	-1	17	186
	Max	12	4	20	
	Average	10	2	18	
Recoveries	Min	40	-106	185	1680
	Max	142	44	239	
	Average	84	-45	210	

et al., [9], Khan et al. [19, 27] used time series models to forecast future’s scenario of COVID-19 in Pakistan. The results of Aslam et al. [9] show increasing trend and suggested eightfold increase in recoveries and an increase of

2.7 times in new cases in the end of May 2020. However, the 10 days ahead forecast of Khan et al. [19] showed that the maximum number of new cases, maximum deaths and maximum recovered cases will be 9,334, 210 and 4,761, respectively, in Pakistan. The results of this study were more realistic as compared to that of Aslam et al. [9] in terms of number of cases. The findings of Carcione et al. [6] suggested that the ending phase of COVID-19 would be in the middle of August 2020 in the Italian region of Lombardy. Fanneh et al. [28] forecasted infections, fatalities and recoveries about COVID-19 in sub-Saharan region of Africa and concluded that the infections number will be 141,733 and 986,059 in the next 30 and 60 days, respectively. The results of the current study can provide valuable information on infections, fatalities and recoveries about COVID-19 in Algeria. Therefore, it is suggested to concerned departments and agencies to consider the findings of this study as guidelines while developing future’s plans regarding COVID-19. It is to mention that this study is adapted only for short-term forecasting and based only on official reported numbers knowing that the real numbers of cases are much higher. At last one of the limitation of our model is that it does not include preventive measures and consider them as stable in the forecasted period.



6 Conclusions

Time series VAR model has been used to describe the data generating process of COVID-19 (daily new infections, fatalities and recoveries) in Algeria. VAR model was used for 20 days ahead forecast and it was noted that the forecast results of all these variables follow the trend of observed data reasonably. For daily new infections, it was noted that the minimum number of new cases will be 63 [CI = -89 to 214] in the upcoming 20 days. The maximum number of new cases will be 147 [CI = 108–186]. The minimum number of forecasted fatalities will be 8 [CI = -1 to 17] while maximum number will be 12 [CI = 4–20]. The forecast results suggest also that there will be 40 [CI = -106 to 185] and 142 [CI = 4–239] minimum and maximum number of recovered cases, respectively. The results suggest that the average number of recoveries is less than the average number of infections, and therefore, probably the number of infected people will increase in future. The findings of this study may be helpful for health, education and other concerned departments in Algeria.

Abbreviations

ARIMA: Autoregressive integrated moving average; COVID-19: Coronavirus disease 2019; LSTM: Long short-term memory networks; SIR: Susceptible–infectious–recovered; SEIR: Susceptible–exposed–infectious–recovered; TSMs: Time series models; VAR: Vector autoregressive.

Acknowledgements

We would thank Dr OUABDESSELAM Lyes from the CRAPC Bousmail (Algeria) for his encouragements and his valuable help.

Authors' contributions

ML conceived the research, collected the data and wrote the manuscript. FK analyzed the data and wrote the manuscript. ML and FK revised and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), H-12 Sector, Islamabad 44000, Pakistan. ²Department of Agro-Veterinary Science, Faculty of Natural and Life Sciences, University of Ziane Achour, Road of Moudjbara, BP 3117, 17000 Djelfa, Algeria.

Received: 12 February 2021 Accepted: 8 August 2021

Published online: 19 August 2021

References

- Issanov A, Amanbek Y, Abbay A, Adambekov S, Aljofan M, Kashkynbayev A, Gaipov A (2020) COVID-19 outbreak in post-soviet states: modeling the best and worst possible scenarios. *Electron J Gen Med* 17(6):em256. <https://doi.org/10.29333/ejgm/8346>
- <https://www.who.int/news/item/29-06-2020-covid-timeline>. Accessed 1 Aug 2021
- Johns Hopkins University of Medicine, Coronavirus resource center: <https://coronavirus.jhu.edu/map.html>. Accessed 31 Aug 2020.
- Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv preprint arXiv: 2002.06563, 2020.
- Alsayed A, Sadir H, Kamil R, Sari H (2020) prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia, 2020. *Int J Environ Res Public Health* 17:4076. <https://doi.org/10.3390/ijerph17114076>
- Carcione JM, Santos JE, Bagaini C, Ba J (2020) A simulation of a COVID-19 epidemic based on a deterministic SEIR model. *Front Public Health* 8:230. <https://doi.org/10.3389/fpubh.2020.00230>
- Bagal DK, Rath A, Barua A, Patnaik D (2020) Estimating the parameters of the susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods. *Chaos Solitons Fractals*. <https://doi.org/10.1016/j.chaos.2020.110154>
- Gordeev D, Singer P, Michailidis M, Müller M, Ambati S. Backtesting the predictability of COVID-19. arXiv: 2007.11411v1 [physics.soc-ph] 22 Jul 2020.
- Aslam M (2020) Using the Kalman filter with Arima for the COVID-19 pandemic dataset of Pakistan. *Data Brief* 31:105854. <https://doi.org/10.1016/j.dib.2020.105854>
- Bentout S, Chekroun A, Kuniya T (2020) Parameter estimation and prediction for coronavirus disease outbreak 2019 (COVID-19) in Algeria. *AIMS Public Health* 7(2):306–318. <https://doi.org/10.3934/publichealth.2020026>
- Anne WR (2020) ARIMA modelling of predicting COVID-19 infections. medRxiv. <https://doi.org/10.1101/2020.04.18.20070631>
- Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, Nanda C, Sharma S, Sharma YD, Rabaaan AA, et al (2020) Prediction of the covid-19 pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (arima) model. *JMIR Public Health Surveill* 6(2):e19115. <https://doi.org/10.2196/19115>
- Pourghasemi HR, Pouyan S, Farajzadeh Z, Sadhasivam N, Heidari B, Babaei S, Tiefenbacher JP (2020) Assessment of the outbreak risk, mapping and infestation behavior of COVID-19: application of the autoregressive integrated and moving average (ARIMA) and polynomial models. *PLoS One* 15(7):e0236238. <https://doi.org/10.1371/journal.pone.0236238>
- Maleki M, Mahmoudi MR, Wraith D, Pho KH (2020) Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Med Infect Dis*. <https://doi.org/10.1016/j.tmaid.2020.101742>
- Chimmula VKR, Zhang L (2020) Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fract* 135:109864. <https://doi.org/10.1016/j.chaos.2020.109864>
- Yonar H, Yonar A, Tekindal MA, 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. *Eur J Med Oncol* 4(2):160–165. <https://doi.org/10.14744/ejmo.2020.28273EJMO>
- Mathevet T, Lepiller M, Mangin A (2004) Application of time series analyses to the hydrological functioning of an Alpine karstic system: the case of Bange-L'Eua-Morte'. *Hydrol Earth Syst Sci* 8(6):1051–1064
- Khan F, Pilz J (2018) Modelling and sensitivity analysis of river flow in the Upper Indus Basin. *Pakistan Int J Water* 12(1):1–21. <https://doi.org/10.1504/IJW.2018.090184>
- Khan F, Saeed A, Ali S (2020) Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan. *Chaos Solitons Fractals* 140:110189. <https://doi.org/10.1016/j.chaos.2020.110189>
- Hannan EJ, Quinn BG (1979) The determination of the order of autoregression. *J R Stat Soc B* 41:190–195
- Akaike H (1970) Statistical predictor identification. *Ann Inst Stat Math* 22:203–217
- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) 2nd international

- symposium on information theory. Akademia Kiado, Budapest, pp 267–281
23. Schwarz G (1978) Estimating the dimension of a model. *Ann Statist* 6:461–464
 24. Baleilar M (2019) Package mFilter. <https://cran.r-project.org/web/packages/mFilter/mFilter.pdf>. Accessed 3 Sept 2020.
 25. Pfaff, B. (2018). Package vars. <https://cran.r-project.org/web/packages/vars/vars.pdf>. Accessed 3 Sept 2020.
 26. Yousaf M, Zahir S, Riaz M, Hussain SM, Shah K (2020) Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. *Chaos Solitons Fractals* 138:109926. <https://doi.org/10.1016/j.chaos.2020.109926>
 27. Khan F, Ali S, Saeed A, Kumar R, Khan AW (2021) Forecasting daily new infections, deaths and recovery cases due to COVID-19 in Pakistan by using Bayesian Dynamic Linear Models. *PLoS ONE* 16(6):e0253367. <https://doi.org/10.1371/journal.pone.0253367>
 28. Fanneh MM, Jobarteh M, Mendy JT, Jawo A (2020) Modeling and forecasting infections, fatalities and recoveries from COVID-19 pandemic in SSA: a case of the 10 hotspot in Sub-Saharan Africa. *Technium Soc Sci J* 9:308–317. <https://doi.org/10.47577/tssj.v9i1.1045>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
