Tak et al. Scr Med 2021 Mar;52(1):6-14. DOI:10.5937/scriptamed52-29893
ORIGINAL ARTICLE

Indian COVID-19 Dynamics: Prediction Using Autoregressive Integrated Moving Average Modelling

Amit Tak,[1] Sunita Dia,[2] Mahendra Dia,[3] Todd C Wehner[3]

Abstract

Background: The forecasting of Coronavirus Disease-19 (COVID-19) dynamics is a centrepiece in evidence-based disease management. Numerous approaches that use mathematical modelling have been used to predict the outcome of the pandemic, including data-driven models, empirical and hybrid models. This study was aimed at prediction of COVID-19 evolution in India using a model based on autoregressive integrated moving average (ARIMA).
Material and Methods: Real-time Indian data of cumulative cases and deaths of COVID-19 was retrieved from the Johns Hopkins dashboard. The dataset from 11 March 2020 to 25 June 2020 (n = 107 time points) was used to fit the autoregressive integrated moving average model. The model with minimum Akaike Information Criteria was used for forecasting. The predicted root mean square error (PreRMSE) and base root mean square error (BaseRMSE) were used to validate the model.
Results: The ARIMA (1,3,2) and ARIMA (3,3,1) model fit best for cumulative cases and deaths, respectively, with minimum Akaike Information Criteria. The prediction of cumulative cases and deaths for next 10 days from 26 June 2020 to 5 July 2020 showed a trend toward continuous increment. The PredRMSE and BaseRMSE of ARIMA (1,3,2) model were 21,137 and 166,330, respectively. Similarly, PredRMSE and BaseRMSE of ARIMA (3,3,1) model were 668.7 and 5,431, respectively.
Conclusion: It is proposed that data on COVID-19 be collected continuously, and that forecasting continue in real time. The COVID-19 forecast assist government in resource optimisation and evidence-based decision making for a subsequent state of affairs.

Keywords: Autoregressive integrated moving average; COVID-19; Epidemic curve; Forecast; Mathematical modelling; Prediction.

(1) ICMR - National Centre for Disease Informatics and Research, Bengaluru, Karnataka, India.
(2) Department of Rheumatology, Medstar Washington Hospital Center, Washington DC 20010, USA.
(3) Department of Horticultural Science, North Carolina State University, Raleigh, NC 27695-7609, USA.

Correspondence:
AMIT TAK
E: dramittak@gmail.com

ARTICLE INFO
Received: 16 December 2020
Revision received: 15 February 2021
Accepted: 15 February 2021

Introduction

Coronavirus Disease-19 (COVID-19), caused by the novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has caused a pandemic with global devastation to human life and health. SARS-CoV-2 is closely related to two bat-derived SARS like coronaviruses, bat-SL-CoVZC45 and bat-SL-CoVZXC21. It is transmitted by human-to-human transmission via droplets or direct contact and the mean incubation period is 6.4 days.[1] According to World Health Organisation, 10,185,374 confirmed cases and 503,862 deaths have been recorded by 1 July 2020.[2] Furthermore, India recorded 568,092 confirmed cases and 17,400 deaths by the same time.[3] The novelty and rapid spread of SARS CoV-2 has challenged medical science across the disciplines of epidemiology, clinical signs and

This article should be cited as follows: Tak A, Dia S, Dia M, Wehner TC. Indian COVID-19 dynamics: prediction using autoregressive integrated moving average modelling. Scr Med 2021 Mar;52(1):6-14.

symptoms, pathophysiology, disease progression and evolution and management and its preventive protocol. Meanwhile, mathematical models have been used to predict the course of disease and mortality.[4-7] Such models have potential in disease management as well as preventive protocols that are cost-effective that can help optimum allocation of resources to manage the disease.[8-14]

An econometric model has been proposed in the present study to predict and extrapolate the transmission of COVID-19. The model makes use of autoregressive integrated moving average modelling on the epidemiological dataset of COVID-19. The objective of this study was to estimate the forecast of COVID-19 cumulative cases and deaths for India. It is proposed that data on COVID-19 should be collected continuously and that forecasting continues in real time to assist governments in evidence-based decision making.

Methods

A descriptive study was run to forecast COVID-19 evolution in the Indian subcontinent. The time series of COVID-19 cumulative cases and deaths from 11 March 2020 to 25 June 2020 (n = 107 time points) was used in prediction using autoregressive integrated moving average modelling. A useful ARIMA model depends on the number of sample time points, and a good series would have more than 50 sample points.[15]

Data acquisition

The Indian data of COVID-19 cumulative cases and deaths from 11 March 2020 to 25 June 2020 were sourced from the official website of Johns Hopkins University Center for Systems Science and Engineering (https://systems.jhu.edu/) and the repository (https://github.com/CSSEGISandData/COVID-19).[16] Excel 2010 was used to build the database.

To validate the model, dataset was distributed into a training set and a validation set. The training dataset from 11 March 2020 to 4 June 2020 (n = 86) was used to fit autoregressive integrated moving average (ARIMA) model and estimation of parameters. The validation dataset from 5 June 2020 to 25 June 2020 was used to validate the model (Table 1).

Table 1: Distribution of complete datasets of cumulative cases and death cases into training and validation subsets

Sub-datasets	Cumulative cases	Death cases
Training set	11 Mar - 04 Jun 2020 (n = 86) (T×C dataset)	11 Mar - 04 Jun 2020 (n = 86) (T×D dataset)
Validation set	05 Jun - 25 Jun 2020 (n = 21) (V×C dataset)	05 Jun - 25 Jun 2020 (n = 21) (V×D dataset)

Model description

The autoregressive integrated moving average model was proposed to estimate the forecast of COVID-19 evolution.[17-19] Box-Jenkins methodology was followed (Figure 1). The methodology encompasses three phases of identification, estimation and testing and application. The Identification phase involves data preparation and model selection. The data were analysed for trends and seasonal components. The Augmented Dickey-Fuller (ADF) unit root test was performed on time series to examine stationarity. Logarithmic transformation and differencing operations were performed to stabilise the time series. Selection of ARIMA model was required to establish the order of the autoregressive (AR) process 'p', the differencing operator 'd', and the order of moving average (MA) process, 'q'. The succeeding system of mathematical equations delineates the ARIMA(p,d,q) model:

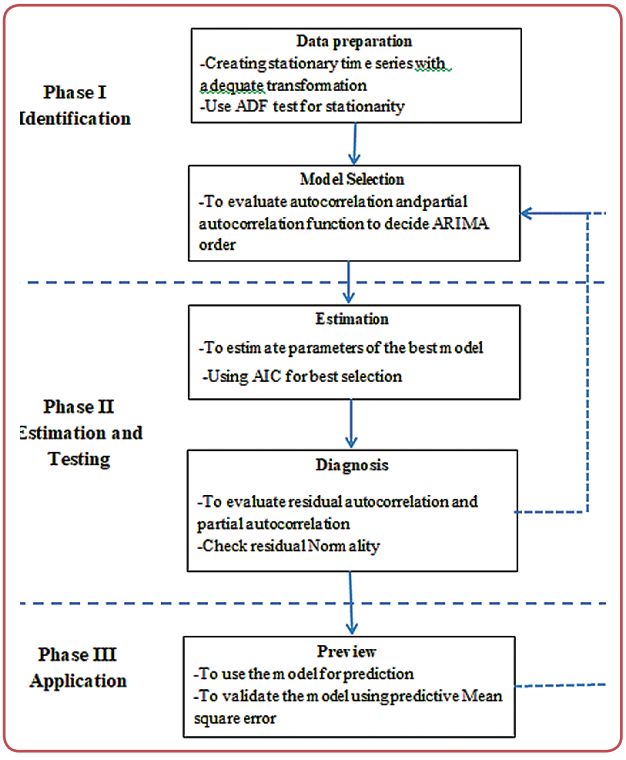


Figure 1: Flowchart outline Box Jenkins Methodology for implementation of autoregressive integrated moving average modeling

$$W_t = \alpha_1 W_{(t-1)} + \alpha_2 W_{(t-2)} + \ldots + \alpha_{(t-p)} W_{(t-p)} + a_t + \beta_1 a_{(t-1)} + \beta_2 a_{(t-2)} + \ldots + \beta_q a_{(t-q)}$$

$$W_t = \nabla^d X_t$$

where $\{X_t\}$ is the original time series, $\{W_t\}$ is the time series acquired after differencing operation on $\{X_t\}$, ∇^d is the difference operator, α_p, β_q are parameters and a_t is the white noise. Since,

$$\nabla^d \equiv (1 - B)^d$$

where B is the backward shift operator,[4] a series $\{X_t\}$ is integrated of order ^{d} if

$$W_t = (1 - B)^d X_t$$

The autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to guide the autoregressive and moving average order, respectively. The Akaike Information Criterion (AIC) was used to select the best model. The AIC considers the maximum log-likelihood estimation (Log L) and number of parameters as the criteria for best model selection. The minimisation of the following equation is required:

$$AIC = -2 \log (\text{maximum likelihood}) + 2k$$

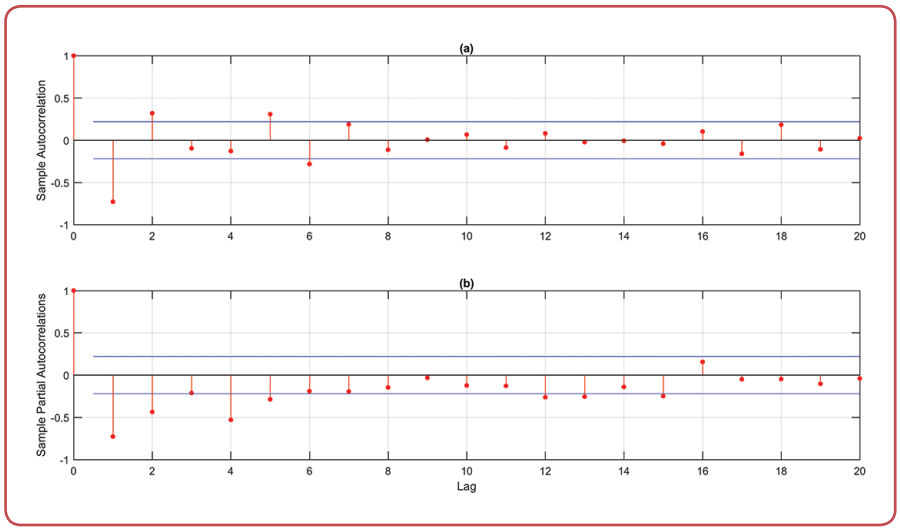
where k = p + q + 1, 2k serves as a penalty function. The model having minimum AIC value was considered the best. Estimation of parameters was accomplished using the method of maximum likelihood estimation. The assumptions of the model were scrutinised with residual diagnostics. The forecasts of cumulative cases and deaths was performed from the best selected ARIMA model. The validation of the model was performed from the validation dataset by enumerating Predicted Root Mean Square Error (PredRMSE) and comparing it with Base Root Mean Square Error (Base RMSE). The forecast of cumulative cases and deaths was performed from 26 Jun to 05 Jul 2020.

Statistical analysis

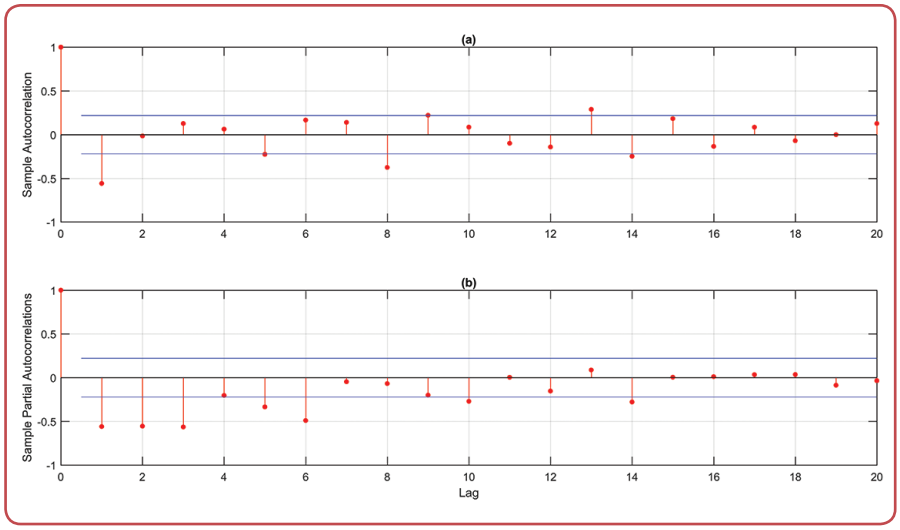
The estimate from the ARIMA model follow a normal distribution, and variation of the forecast was considered within 95 % confidence intervals. The MS Excel 2010 was used to maintain the database and MATLAB 2016a (version 9.0.0.341360) was used for analysis.[20]

Results

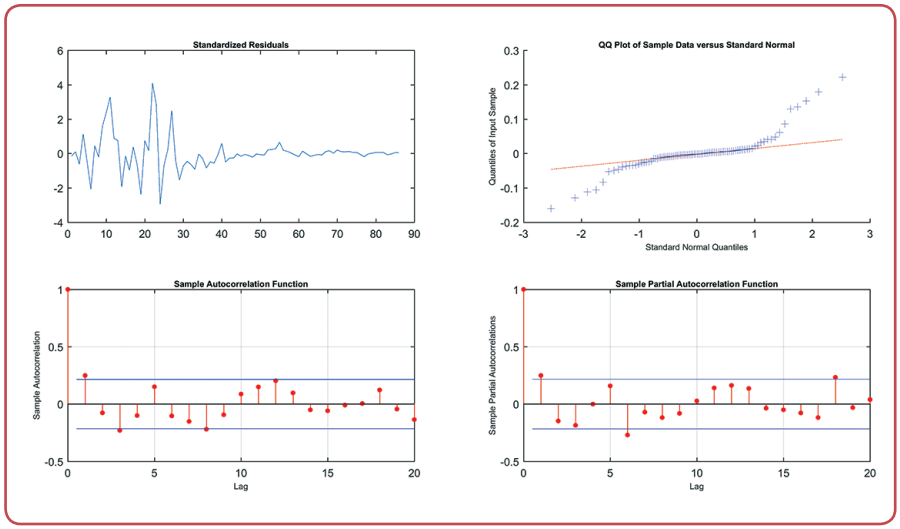
To achieve stationarity, the logarithmic transformation of datasets (cumulative cases and deaths) was performed as they were showing upward trends. Further difference operations were performed on cumulative cases (d = 3) and death (d = 2) datasets. The Augmented Dickey-Fuller (ADF) test showed stationarity (p value < 0.05). The autocorrelation function and partial autocorrelation function were calculated and correlograms were used to guide the autoregressive order and the moving average order of the ARIMA models (Supplemental Figure 1 and 2). The ARIMA



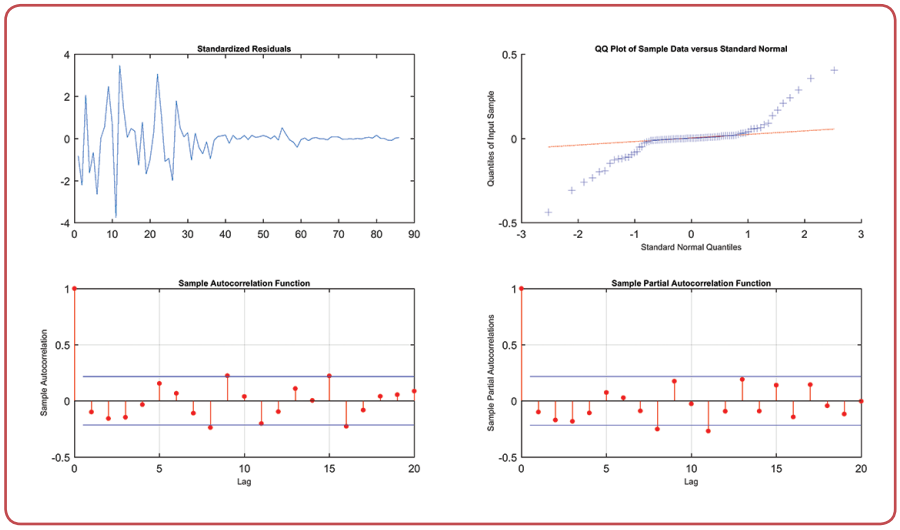
Supplemental Figure S1: Correlograms of time series of cumulative cases of COVID-19 (a) sample autocorrelation function (b) sample partial autocorrelaton function



Supplemental Figure S2: Correlograms of time series of death cases of COVID-19 (a) sample autocorrelation function (b) sample partial autocorrelaton function



Supplemental Figure S3: Line plots, QQ plots and correlograms of residual diagnostic of differenced logarithmic transformed time series of cumulative cases of COVID-19. Line plots showing random distribution of standardised residuals and QQ plot showing straight line in the mid portion and the autocorrelation and partial autocorrelation functions are within random limits, represent normality of residuals



Supplemental Figure S4: Line plots, QQ plots and correlograms of residual diagnostic of differenced logarithmic transformed time series of death cases of COVID-19. Line plots showing random distribution of standardised residuals and QQ plot showing straight line in the Mid-portion and the autocorrelation and partial autocorrelation functions are within random limits, represent normality of residuals

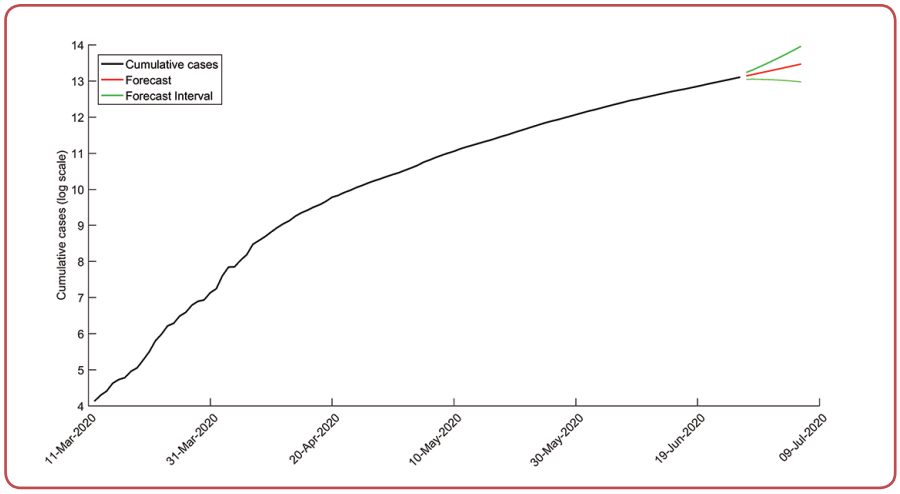


Figure 2: Line plot shows observed (black line plot) and predicted (magenta line plot) cumulative cases of COVID-19 with 95 % confidence interval (green line plot). The forecast is based on ARIMA (1,3,2) model fitted with time series of cumulative cases from 11 March to 25 Jun 2020

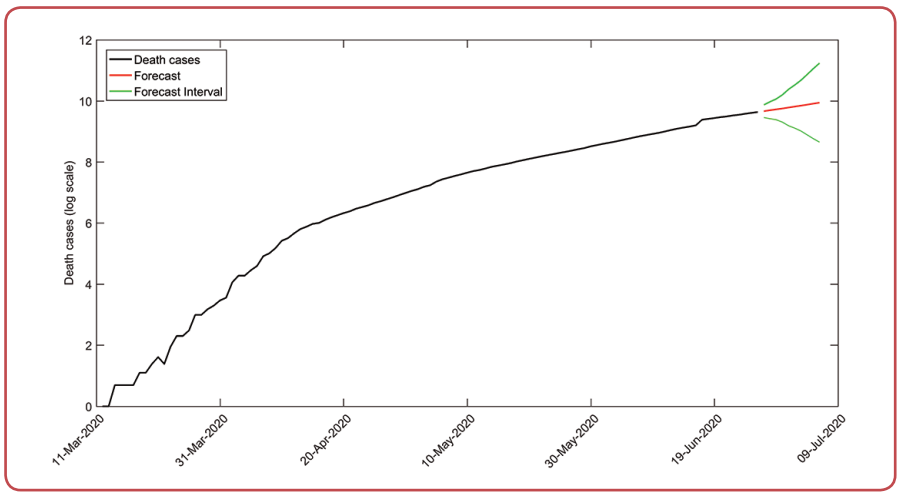


Figure 3: Line plot shows observed (black line plot) and predicted (magenta line plot) death cases of COVID-19 with 95 % confidence interval (green line plot). The forecast is based on ARIMA (3,3,1) model fitted with time series of death cases from 11 March to 25 Jun 2020

Supplemental Table S1: Parameters, errors and statistics of ARIMA (1,3,2) model, fitted from training dataset of cumulative cases of COVID-19

Parameter	Value	Error	Statistics
Constant	0.000	0.0002	0.171
AR {1}	-0.398	0.036	-10.923
MA {1}	-1.756	0.048	-36.406
MA {2}	0.756	0.047	16.092
Variance	0.003	0.0002	10.190

Supplemental Table S2: Parameters, errors and statistics of ARIMA (3,3,1) model, fitted from training dataset of death cases of COVID-19

Parameter	Value	Error	Statistics
Constant	0.0002	0.001	0.158
AR {1}	-1.118	0.052	-21.502
AR {2}	-1.078	0.058	-18.482
AR {3}	-0.606	0.053	-11.430
MA {1}	-1	0.064	-15.525
Variance	0.013	0.001	9.700

Table 2: Goodness of fitted models for cumulative cases and deaths cases using prediction root mean square error and base root mean square error

Dataset	Model	PredRMSE	BaseRMSE
Cumulative cases	ARIMA (1,3,2)	21137	166330
Death cases	ARIMA (3,3,1)	668.70	5431

**ARIMA: autoregressive integrated moving average; PredRMSE: predicted root mean square error; BaseRMSE: base root mean square error*

(1,3,2) and ARIMA (3,3,1) model for cumulative cases (AIC = 2.32) and death cases (AIC = 108.75) was chosen respectively with the lowest AIC values. Goodness of fit and model assumptions were tested with residual analysis. The graph of standardised residuals showed random distribution of residuals. The middle values of the QQ plot were on a straight line. The autocorrelation function and partial autocorrelation functions within random limits showed normal distribution and independence (Supplemental Figure 3 and 4). The

Table 3: Forecast of cumulative cases and death cases using ARIMA (1,3,2) and ARIMA (3,3,1) models respectively. The dataset used for prediction is from 11 Mar 2020 to 25 Jun 2020 (n = 107 time points)

Date	Cumulative cases			Cumulative deaths		
	Forecast	95 % CI		Forecast	95 % CI	
		LL	UL		LL	UL
26 Jun 20	508210	378270	638150	15732	12399	19065
27 Jun 20	526760	396820	656700	16243	12910	19575
28 Jun 20	546010	416070	675950	16711	13379	20044
29 Jun 20	566050	436100	695990	17198	13865	20530
30 Jun 20	586900	456960	716840	17740	14408	21073
1 Jul 20	608630	478690	738570	18314	14982	21647
2 Jul 20	631290	501350	761230	18873	15540	22205
3 Jul 20	654940	525000	784890	19480	16147	22813
4 Jul 20	679660	549710	809600	20141	16808	23474
5 Jul 20	705500	575560	835440	20818	17485	24150

**CI: confidence interval; LL: lower limit of CI; UL: upper limit of CI*

parameters of both the ARIMA models were estimated (Supplemental Table S1 and S2).

Model validation was conducted by detecting the differences between the observed values and predicted values from the validation dataset. The PredRMSE and BaseRMSE of ARIMA (1,3,2) model for time series of cumulative cases was 21137 and 166330, respectively. Similarly, PredRMSE and BaseRMSE of ARIMA (3,3,1) model for time series of death cases were 668.7 and 5431, respectively. Lower values of PredRMSE than Base RMSE indicated a good fit (Table 2). The estimation of forecast for cumulative cases and death cases were performed at 95 % confidence interval (Figures 2 and 3; Table 3).

Discussion

The dynamics of any infectious disease involve interactions of three elements - host, agent and environment.[21] India lies in both the Northern and Eastern hemispheres, with latitudes to north and longitudes to east results in high environmental variability, and subsequent effects on human behaviour and society.[22] The mathematical models are a centrepiece to forecasting the dynamics of infectious disease pandemics. The fundamental models used in prediction include data-driven,[23-27] empirical,[28-30] hybrid[31-32] models. The parameters of empirical models (incubation period, attack rate, and recovery rate) possess probability distributions such as Ehrlang and the Poisson distribution.[21] The data-driven models include ARIMA, single-input and single-output (SISO) models[33] and AI-based model (machine learning and deep learning techniques).[24] The significance of ARIMA model lies in modelling nonstationary time series. The ARIMA model assumes the trends will continue in the future indefinitely as against the empirical model which assume convergence. Few studies used nonparametric models like Fourier decomposition methods to predict turn-around dates of the epidemic and the results were found to agree with popular SIR models.[34]

SARS-CoV-2 expressed its presence in India with the first case diagnosed on 30 January 2020. On 1 July 2020, the cumulative cases and deaths in India reached 568,092, and 17,400, respectively.[3] India ranked 57th among the 100 countries in Global Health Security Index 2019, a scale to gauge preparedness for the outbreak of serious infectious diseases.[35] The forecast of COVID-19 cases helps government agencies in early preparedness to combat subsequent state of affairs. The patients of COVID-19 can be classified into confirmed cases, recovered cases, admitted, and death cases. All categories have differential importance from the management point of view. The requirement of hospital beds, medical equipment, hospital staff is a function of the number of admitted cases. Similarly, procurement of plasma for antibody therapy is a function of the number of recovered cases.

The present study forecasts the cumulative cases and deaths for India from 26 June 2020 to 5 July 2020. The forecast shows continuously increasing trends for both cumulative cases and deaths and shows no decreasing trends until October 2020.

The ARIMA model was proposed by various authors for forecasting the COVID-19 evolution. The data from 10 January 2020 to 20 February 2020 was used to fit ARIMA (1,0,4) and ARIMA (1,0,3) models for cumulative diagnosis and newer diagnoses to forecast the next two days.[36] The ARIMA and wavelet hybrid model was proposed to forecast ten time points of cumulative cases from 05 Apr 2020 to 14 Apr 2020 for India. The forecast showed oscillations may be due to effects of lockdown.[32] In another data-driven model, using bidirectional LSTM (long short-term memory) model, 15 days prediction of actual cases in India from 30 April 2020 to 14 May 2020 showed an error of less

than 3 %.[31] Susceptible-Exposed-Infectious-Recovered (SEIR) model was used to predict cumulative cases of COVID-19 in India during lockdown and post lockdown. The model predicts the peak of cumulative cases around 43,000 in mid-May. However, 7-21 % increase in peak value of cases was predicted for post lockdown period, reflecting relaxation in control strategies.[30] The evolution of COVID-19 in topmost affected states of India was done using SISO model.[33] The most severely affected states were Maharashtra, Gujarat, Tamil Nadu, Delhi, and Rajasthan.[33,37-39] One time series analysis used genetic programming to predict the COVID-19 evolution in India. On 13 May 2020, the cumulative cases and death cases were 80,000 and 2500, respectively. The prediction for the next ten days was done with 142,000 and 4,200 cumulative cases and deaths, respectively, on 23 May 2020.[40] Sujath used linear regression, multilayer perceptron and vector autoregressive method and 80 time points till 10 April 2020, to predict confirmed cases, deaths, and recovered cases from 11 April 2020 to 18 June 2020. Although prediction varies across the methods and did not seem very accurate.[41] Yadav used six regression analysis-based machine learning models for prediction and found six-degree polynomial models predict very close to observed data.[42]

As the epidemic continues, the effect of different interventional strategies inherited in the time series, and thus data-driven model seem to be more accurate than empirical models. The real-time data modelling has been proposed for monitoring trends of the Indian subcontinent. Evidence-based interventions should be implemented to control the pandemic.

Conclusion

The present study produces encouraging results with the potential to serve as a good adjuvant to existing models for continuous predictive monitoring of the COVID-19 pandemic. The forecast of COVID-19 may assist public health authorities and governmental agencies for early preparedness and evidence-based decision making.

The limitations of the study

Firstly, ARIMA model used in forecasting does not capture the non-linear and chaotic dynamics of the pandemic. Secondly, the parameter selection procedure requires repetition with each time series update.

Contribution of Authors

All authors contributed equally.

Competing Interest

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

Data Availability of

Data is available on reasonable request from the corresponding author.

Ethics Declarations

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by Ethical Committee of the SMS Medical College, Jaipur (No.430/MC/EC/2020 dated 26 Jun 2020) and individual consent for this retrospective analysis was waived.

Acknowledgements

None.

Conflict of interest

None.

Abbreviations

ACF:	Autocorrelation function
ADF test:	Augemented Dickey Fuller unit root test
AIC:	Akaike information criterion
ARIMA:	Autoregressive integrated moving average
BaseRMSE:	Base root mean square error
COVID-19:	Coronavirus disease-2019
PACF:	Partial autocorrelation function
PredRMSE:	Prediction root mean square error
SARS CoV-2:	Severe Acute Respiratory Syndrome Coronavirus 2
T × C dataset:	training dataset of cumulative cases of COVID-19
T × D dataset:	training dataset of death cases of COVID-19

References

1. Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. Int J Antimicrob Agents [Internet]. 2020 Mar;55(3):105924. Available from: http://dx.doi.org/10.1016/j.ijantimicag.2020.105924.
2. World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard, cited on: 01 Jul 2020; Available from: https://covid19.who.int/.
3. Ministry of Health and Family Welfare, Government of India, COVID-19 India; cited on: 01 Jul 2020; Available from: https://www.mohfw.gov.in/.
4. Bhandari S, Shaktawat AS, Tak A, Patel B, Shukla J, Singhal S, et al. Logistic regression analysis to predict mortality risk in COVID-19 patients from routine hematologic parameters. Ibnosina J Med Biomed Sci [serial online] 2020 [cited 2020 Jul 1];12:123-9. Available from: http://www.ijmbs.org/text.asp?2020/12/2/123/288204.
5. Bhandari S, Tak A, Singhal S, Shukla J, Patel B, Shaktawat AS, et al. Patient flow dynamics in hospital systems during times of COVID-19: Cox proportional hazard regression analysis. Front Public Health [Internet]. 2020. Available from: doi: 10.3389/fpubh.2020.585850.
6. Bhandari S, Shaktawat AS, Tak A, Shukla J, Gupta J, Patel B, et al. Relationship between ABO blood group phenotypes and nCOVID-19 susceptibility – a retrospective observational study. Scr Med 2020;51(4):217-22. Available from: https://aseestant.ceon.rs/index.php/scriptamed/article/view/29692.
7. Bhandari S, Shaktawat AS, Tak A, Shukla J, Patel B, Singhal S, et al. Evidence-based decision making and covid-19: what a posteriori probability distributions speak. JIDHealth [Internet]. 2020 Dec 31;3(Special 2):286-92. Available from: https://jidhealth.com/index.php/jidhealth/article/view/88.
8. Kakkar S, Bhandari S, Singh A, Sharma R, Mehta S, Gupta J, et al. Coronavirus disease of 2019: The premise for framing strategies towards infection prevention control management. Current Medical Issues [Internet]. 2020;18(3):199. Available from: http://dx.doi.org/10.4103/cmi.cmi_82_20.
9. Bhandari S, Singh A, Sharma R, Rankawat G, Banerjee S, Gupta V, Dube A, Kakkar S, Sharma S, Keswani P, Agrawal A. Characteristics, Treatment Outcomes and Role of Hydroxychloroquine among 522 COVID-19 hospitalized patients in Jaipur City: An Epidemio-Clinical Study. J Assoc Physicians India 2020:13-9.
10. Rankawat G, Bhandari S, Singh A, Bagarhatta M, Dube A, Kakkar S, et al. Evaluation of clinico–Radiological profile and correlation with ultrasonography of the chest in coronavirus disease 2019 pneumonia. Indian J Med Spec [Internet]. 2020;11(2):70. Available from: http://dx.doi.org/10.4103/INJMS.INJMS_55_20.
11. Bhandari S, Singh A, Sharma R, Mehta S, Dube A, Gupta J, Gupta K, Tak A, Kakkar S. The proposed bridging management protocol for COVID-19. Menoufia Med J [serial online] 2020 [cited 2020 Nov 22];33:1109-10. Available from: http://www.mmj.eg.net/text.asp?2020/33/3/1109/296653.
12. Bhandari S, Shaktawat A, Patel B, Dube A, Kakkar S, Tak A, et al. The sequel to COVID-19: the antithesis to life. JIDHealth [Internet]. 1Oct.2020 [cited 22Nov.2020];3(Special1):205-12. Available from: https://www.jidhealth.com/index.php/jidhealth/article/view/69.
13. Bhandari S, Shaktawat AS, Patel B, Rankawat G, Tak A, Gupta JK, et al. Hydroxychloroquine in rheumatological disorders: the potential buffer against coronavirus disease-19? J Med Sci Health 2020;6(3):58-65.
14. Tak A, Das B, Shah M, Dia S, Dia M, Gahlot S. COVID-19 and lockdown in India: evaluation using analysis of covariance. J Antivir Antiretrovir [Internet]. 2021;13:216. Available from: https://www.longdom.org/abstract/covid19-and-lockdown-in-india-evaluation-using-analysis-of-covariance-64008.html.
15. Box GEP, Tiao GC. Intervention Analysis with Applications to Economic and Environmental Problems. J Am Stat Assoc [Internet]. 1975 Mar;70(349):70–9. Available from: http://dx.doi.org/10.1080/01621459.1975.10480264.
16. Johns Hopkins University Center for Systems Science and Engineering, 2019. (Accessed: 25th Jun 2020). Available from: https://github.com/CSSEGISandData/COVID-19
17. Cryer JD, Chan KS. Models for non-stationary time series in time series analysis: with applications in R. 2nd edition. Berlin/Heidelberg: Springer Science+Business Media; 2008. p. 98-99.
18. Indrayan, Abhaya, Rajeev Kumar Malhotra relation-

ships: quantitative outcome in medical biostatistics. 4th edition. Boca Raton (FL), USA: CRC Press; 2018. p. 456.
19. Metcalfe AV, Cowpertwait PS. Non-stationary models in introductory time series with R. New York: Springer-Verlag; 2009. p. 137-140.
20. MATLAB Team, Statistics and Machine Learning Toolbox 10.2. version 9.0.0.341360 (R 2016a). Natick, Massachusetts: The Mathworks Inc.
21. Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Hassanien AE. Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Computer Sci [Internet]. 2020 Jun 11;1(4). Available from: http://dx.doi.org/10.1007/s42979-020-00209-9.
22. National Portal of India (cited on: 27 April 2020). Available from: https://www.india.gov.in/india-glance/profile.
23. Yang C, Wang J. A mathematical model for the novel coronavirus epidemic in Wuhan, China. Math Biosci Eng [Internet]. 2020;17(3):2708–24. Available from: http://dx.doi.org/10.3934/mbe.2020148.
24. Chatterjee A, Gerdes MW, Martinez SG. Statistical explorations and univariate timeseries analysis on COVID-19 datasets to understand the trend of disease spreading and death. Sensors [Internet]. 2020 May 29;20(11):3089. Available from: http://dx.doi.org/10.3390/s20113089.
25. Tiwari S, Kumar S, Guleria K. Outbreak trends of coronavirus disease–2019 in India: a prediction. Disaster Med Public Health Prep [Internet]. 2020 Apr 22;1–6. Available from: http://dx.doi.org/10.1017/dmp.2020.115.
26. Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Sci Total Environ [Internet]. 2020 Aug;728:138762. Available from: http://dx.doi.org/10.1016/j.scitotenv.2020.138762.
27. Tuli S, Tuli S, Tuli R, Gill SS. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things [Internet]. 2020 Sep;11:100222. Available from: http://dx.doi.org/10.1016/j.iot.2020.100222.
28. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med [Internet]. 2020:1-6. Available from: https://moh-it.pure.elsevier.com/en/publications/modelling-the-covid-19-epidemic-and-implementation-of-population.
29. Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar TK. A model based study on the dynamics of COVID-19: Prediction and control. Chaos Soliton Fract [Internet]. 2020 Jul;136:109889. Available from: http://dx.doi.org/10.1016/j.chaos.2020.109889.
30. Pai C, Bhaskar A, Rawoot V. Investigating the dynamics of COVID-19 pandemic in India under lockdown. Chaos Soliton Fract [Internet]. 2020 Sep;138:109988. Available from: http://dx.doi.org/10.1016/j.chaos.2020.109988.
31. Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India. Chaos Soliton Fract [Internet]. 2020 Oct;139:110017. Available from: http://dx.doi.org/10.1016/j.chaos.2020.110017.
32. Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. Chaos Soliton Fract [Internet]. 2020 Jun;135:109850. Available from: http://dx.doi.org/10.1016/j.chaos.2020.109850.
33. Rafiq D, Suhail SA, Bazaz MA. Evaluation and prediction of COVID-19 in India: A case study of worst hit states. Chaos Soliton Fract [Internet]. 2020 Oct;139:110014. Available from: http://dx.doi.org/10.1016/j.chaos.2020.110014.
34. Singhal A, Singh P, Lall B, Joshi SD. Modeling and prediction of COVID-19 pandemic using Gaussian mixture model. Chaos Soliton Fract [Internet]. 2020 Sep;138:110023. Available from: http://dx.doi.org/10.1016/j.chaos.2020.110023.
35. Tabish SA. The COVID-19 pandemic: Emerging perspectives and future trends. J Public Health Res [Internet]. 2020 Jun 4;9(1). Available from: http://dx.doi.org/10.4081/jphr.2020.1786.
36. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data Brief [Internet]. 2020 Apr;29:105340. Available from: http://dx.doi.org/10.1016/j.dib.2020.105340.
37. Bhandari S, Shaktawat AS, Tak A, Patel B, Gupta K, Gupta J, et al. A multistate ecological study comparing evolution of cumulative cases (trends) in top eight COVID-19 hit Indian states with regression modeling. Int J Acad Med [serial online] 2020 [cited 2020 Jul 1];6:91-5. Available from: http://www.ijam-web.org/text.asp?2020/6/2/91/287965.
38. Kakkar S, Bhandari S, Shaktawat A, Sharma R, Dube A, Banerjee S, et al. A preliminary clinico-epidemiological portrayal of COVID-19 pandemic at a premier medical institution of North India. Ann Thorac Med [Internet]. 2020;15(3):146. Available from: http://dx.doi.org/10.4103/atm.ATM_182_20.
39. Bhandari S, Sharma R, Singh Shaktawat A, Banerjee S, Patel B, Tak A, et al. COVID-19 related mortality profile at a tertiary care centre: a descriptive study. Scr Med 2020;51(2):69-73. DOI:10.5937/scriptamed51-27126.
40. Salgotra R, Gandomi M, Gandomi AH. Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming Chaos Soliton Fract [Internet]. 2020 Sep;138:109945. Available from: http://dx.doi.org/10.1016/j.chaos.2020.109945.
41. Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. Stoch Env Res Risk A [Internet]. 2020 May 30;34(7):959–72. Available from: http://dx.doi.org/10.1007/s00477-020-01827-8.
42. Yadav RS. Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India. International Journal of Information Technology [Internet]. 2020 May 26; Available from: http://dx.doi.org/10.1007/s41870-020-00484-y.