ARTICLE IN PRESS

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Analysis of 'earlyR' epidemic model and time series model for prediction of COVID-19 registered cases

Karthick Kanagarathinam [a,*], Ebrahem A. Algehyne [b], Kavaskar Sekar [c]

[a] Department of EEE, GMR Institute of Technology, Rajam, Andhra Pradesh, India
[b] Department of Mathematics, Faculty of Sciences, University of Tabuk, Saudi Arabia
[c] Department of EEE, Panimalar Engineering College, Chennai, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 24 September 2020
Accepted 4 October 2020
Available online xxxx

Keywords:
Early prediction of epidemic
Forecasting
ARIMA Model
Time series analysis
'earlyR' epidemic model

ABSTRACT

The COVID-19 is an epidemic that causes respiratory infection. The forecasted data will help the policy makers to take precautionary measures and to control the epidemic spread. The two models were adopted for forecasting the daily newly registered cases of COVID-19 namely 'earlyR' epidemic model and ARIMA model. In earlyR epidemic model, the reported values of serial interval of COVID-19 with gamma distribution have been used to estimate the value of R_0 and 'projections' package is used to obtain epidemic trajectories by fitting the existing COVID-19 India data, serial interval distribution, and obtained R0 value of respective states. The ARIMA model is developed by using the 'auto.arima' function to evaluate the values of (p, d, q) and 'forecast' package is used to predict the new infected cases. The methodology evaluation shows that ARIMA model gives the better accuracy compared to earlyR epidemic model.

Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

1. Introduction

The republic of India comprises of 28 states and 8 union territories. The COVID-19 has spread rapidly worldwide and reported 82,42,999 cases total confirmed cases and 4,45,535 deaths globally and India has 3,66,946 as total confirmed cases and 12,237 deaths as on 18th June 2020 [4]. The COVID-19 is zoonotic that are initially suspected in Wuhan, China region with 27 cases of unknown pneumonia. The general signs of infection include fever, cough, respiratory symptoms and breathing difficulties. The medical student from Wuhan, China is the first case in India and was diagnosed on 30th January 2020 at Thrissur, Kerala state. The first case of Tamil Nadu is reported from Kancheepuram, a 45 years old male with travel history to Oman, arrived on 28th February 2020. He was hospitalized on 4th March 2020 and diagnosed on 7th March 2020. The state of Maharashtra is a hotspot of COVID-19 in India. The first symptomatic case of Maharashtra is reported from Pune on 8th March 2020 who had travel history to Dubai and diagnoses confirmed on 10th March 2020 [3]. The government has announced nationwide lockdown for 21 days (i.e, upto 14th April 2020) on 24th March 2020 as one of the precautionary measure against COVID-19 epidemic spread and now phase-5 lockdown is extended up to 30th June 2020 with certain relaxations. The accurate early prediction epidemic model would advance the prevention and control measures. The vaccination for COVID-19 is in the pipeline and there are no explicit effective drug combinations or antiviral with high-level evidence [10]. The premature relaxation may result second wave of infection [11]. Epidemic forecasting models gained huge attention in the field of epidemiology and health sciences. The susceptible-infected-removed (SIR) model [6] is one of the conventional epidemic forecasting models. Zhong, L et al. [14] developed a simple epidemic model for the early prediction of COVID-19 in the Mainland China. Their model depicts the propagation of the SARS in 2003 that has different characteristics from COVID-19. Kenji M and Gerardo C [8] developed a better COVID-19 outbreak model and made projections using sing a MCMC method. Alsharif et al. [1] has developed ARIMA based time series forecasting model to predict the daily and monthly global solar radiation and they compared the results with Monte Carlo method. In this article, 'earlyR' epidemic model and ARIMA model are evaluated to forecast the COVID-19 daily new incidence. It is found that ARIMA model offers better accuracy compared to 'earlyR' epidemic model. The R programming is used for the entire work.

* Corresponding author.
E-mail addresses: kkarthiks@gmail.com, karthick.k@gmrit.edu.in (K. Kanagarathinam), e.algehyne@ut.edu.sa (E.A. Algehyne).

https://doi.org/10.1016/j.matpr.2020.10.086
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

Please cite this article as: K. Kanagarathinam, E.A. Algehyne and K. Sekar, Analysis of 'earlyR' epidemic model and time series model for prediction of COVID-19 registered cases, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.10.086

Nomenclature

ACF	Autocorrelation function
ARIMA	Autoregressive integrated moving average
CI	Confidence Interval
COVID-19	2019 novel coronavirus disease
MAE	Mean absolute error
MAPE	Median absolute percentage error
MASE	Mean absolute scaled error
MCMC	Monte Carlo Markov Chain
ME	Margin of error
MPE	Mean percentage error
PACF	Partial Autocorrelation function
R_0	Reproduction number
RMSE	Root mean square error
WHO	World health organization

2. Methodology

2.1. Data

The COVID-19 data of Maharashtra and Tamil Nadu states has been extracted from the official website [5]. The COVID-19 data of India from 3rd May 2020 to 24th May 2020 and Tamil Nadu data from 07th March 2020 to 16th April 2020 are utilized for methodology evaluation. All the utilized data are available in Annexure I. Fig. 1 shows the daily new incidence of COVID-19 of Maharashtra and Tamil Nadu. The epidemic curve shows for the period from 30th January 2020 to 30th July 2020.

The 'earlyR' is a package available in R programming. The following sections 2.2 and 2.3 explain about the model development followed by methodology evaluation.

2.2. 'earlyR' epidemic model

The proposed 'earlyR' epidemic model is a simplified version of the model introduced by Anne Cori et al. Anne Cori et al. [2]. The Fig. 2 shows the procedural approach of the model. The daily incidence R_0 of COVID-19 is estimated using 'earlyR' and 'incidence' package. The serial interval distribution is an important parameter in the estimation of the transmissibility or R0 of COVID-19 and same has an impact on accuracy of predicting the cumulative number of new cases in near future that helps to take decisions against the public health crisis.

Li, Q. et al., [9] estimated that the incubation period of COVID-19 is 5.2 days with 95% credible interval with six infector-infectee data and found evidence that human-to-human transmission occurred among close contacts. Hiroshi Nishiura et al. [7] estimated the median serial interval as 4.0 days with 95% credible interval with 28 infector-infectee dataset and limiting the data to most certain pair median serial interval is estimated as 4.6 days for COVID-19 with 95% credible interval. The serial interval distributions i.e., mean and standard value is required to estimate the early stage of R0. The mean and standard deviation are assumed as 4.7 days and 2.9 days respectively from the existing work [7] to obtain Maximum-Likelihood R0 value.

The bootstrap strategy is applied for re-sampling with 1000 times for obtaining likely R0 values. The prediction of cumulative daily incidence for the next 10 days has been computed using the 'projections' package in R [12]. The epidemic trajectories are obtained by fitting the existing COVID-19 India data, serial interval distribution, and obtained R0 value of respective state and union territory. The daily incidence obeys Poisson distribution determined by daily infectiousness and it is denoted as,

$$\lambda(\mathrm{t}) = \sum_{n=1}^{t-1} \mathrm{X_n P(t-n)} \quad (1)$$

Where P (t-n) is vector of probability mass function and X_n is real time incidence at time n. The mean, median, minimum and maximum values can be obtained with 95% CI.

2.3. ARIMA model

ARIMA model is specified with three parameters p, d, q and symbolized as ARIMA (p,d,q). The 'p' is the number of autoregressive parameters in the model; d is the number of non seasonal differences; q is the number of moving average parameters in the model.

The X_t is unknown cases in the series of interest at time t, which is a random variable, a p^{th} order of Autoregressive, AR(p) is defined as follows

$$\mathrm{X_t} - \sum_{k=1}^{p} \Phi_k X_{t-k} = \epsilon_t \quad (2)$$

where ϵ_t is a uncertainty affecting our observations, $\Phi_1, \Phi_2, \cdots\cdots \Phi_p$ are the *AR* coefficients.

The *moving average* component of order q, MA(q) is defined as follows

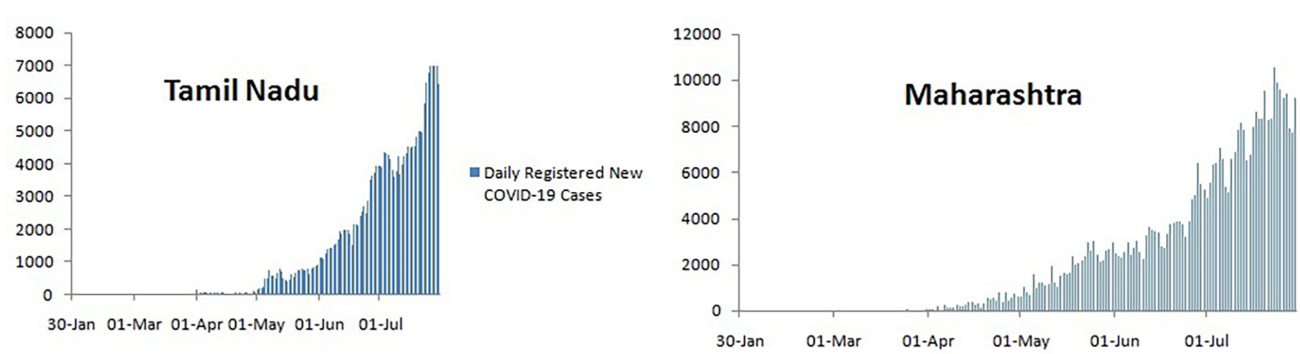

Fig. 1. Daily new incidence of COVID-19.

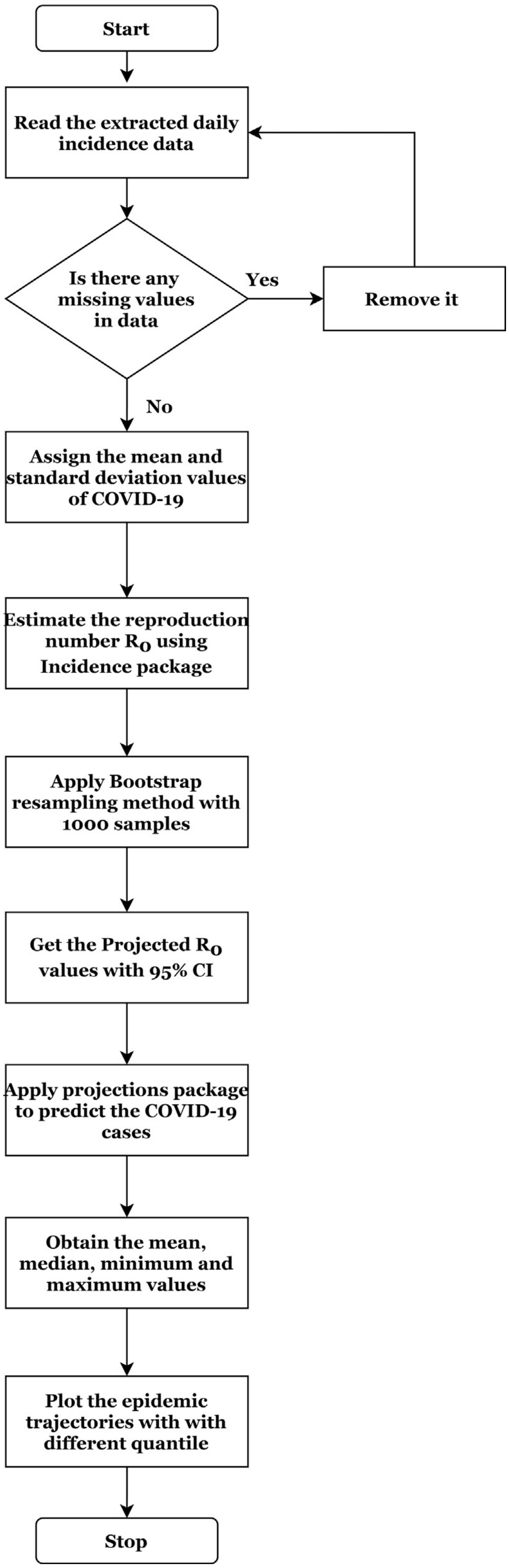

Fig. 2. Flowchart of earlyR epidemic model.

$$X_t = \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} \tag{3}$$

Where $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average parameters.

The degree of differencing d, represents how many cases in the time series are far from being stationary, is defined as follows

$$\Phi(B)\nabla^d X_t = \theta(B)\epsilon_t \tag{4}$$

where $\nabla X_t = (1 - B)X_t$ is the lag 1 differencing operator, and B is the backward shift operator defined by $B^n X_t := X_{t-n}$

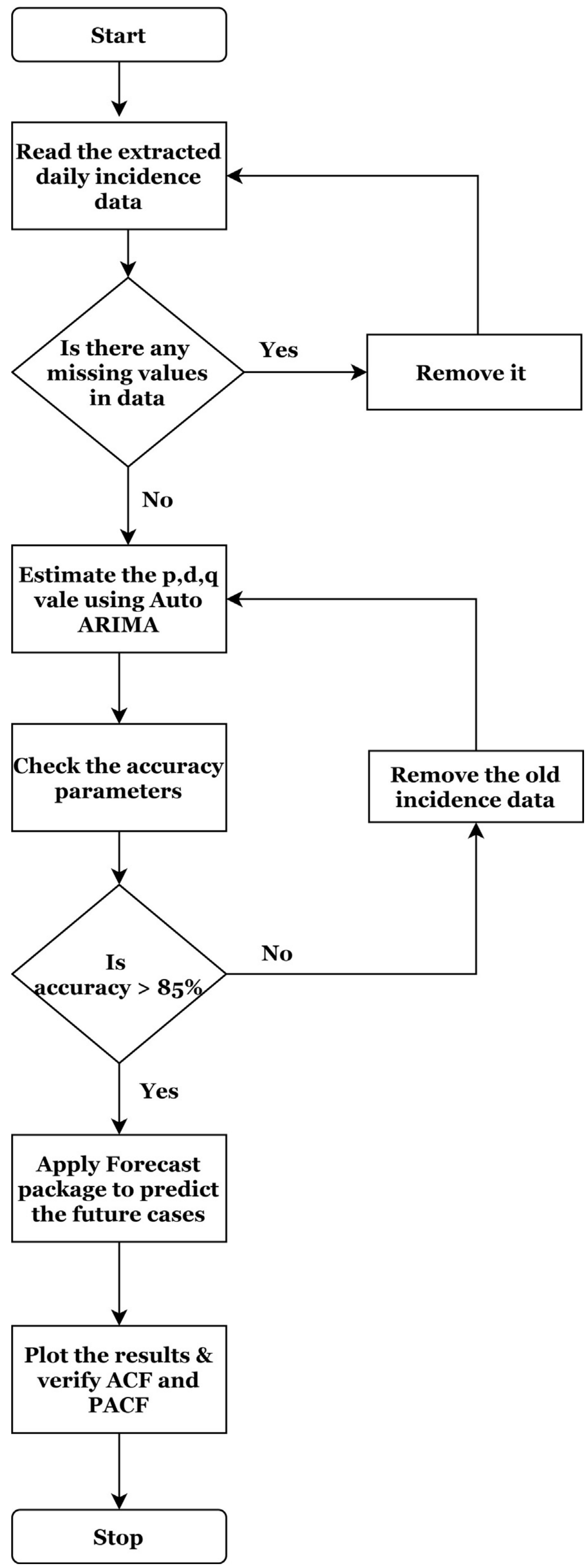

Fig. 3. Flowchart of ARIMA Model.

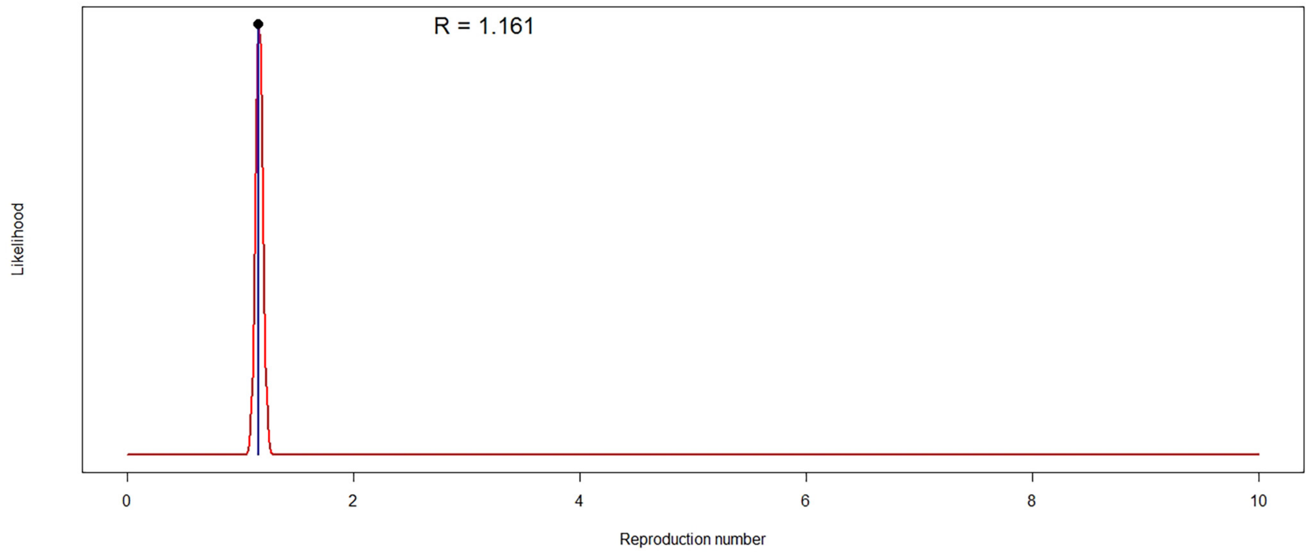

Fig. 4. (a). The likelihood R_0 value of evaluation data of Tamil Nadu (b) Global force of Infection (c). Projected likely R0 values for the next 10 days (d). Predicted incidence with different quantile for the next 10 days.

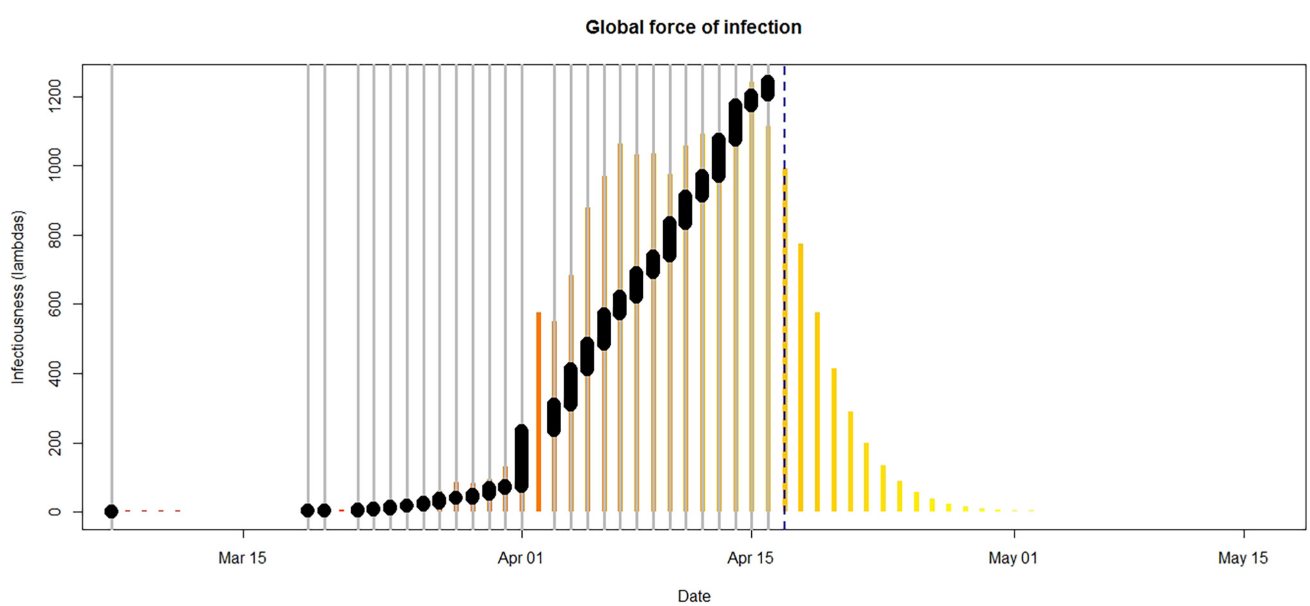

Fig. 4 (*continued*)

Fig. 3 shows the procedural approach of ARIMA model. The specified period data have been recorded in excel file and accessed using 'readxl' package. The missing data should be removed from the series. 'AUTOARIMA' package is applied to estimate the values of (p, d, q). The 'forecast' package is used to obtain the predicted new cases of COVID-19 in India for the next 23 days during evaluation The accuracy of the model has been evaluated using MAPE parameter as (100-MAPE) also checked with ACF and PACF.

3. Methodology evaluation

3.1. Evaluation of 'earlyR' epidemic model with real time data

The Tamil Nadu COVID-19 data from 07th March 2020 to 16th April 2020 is considered as input to the model and it is available in Annexure I. The mean and standard deviation are assumed as 4.7 days and 2.9 days respectively from the existing work [7]. The likelihood R_0 value of the input COVID-19 data for the mentioned period is 1.161 as shown in Fig. 4(a). The global force of infection can be found in Fig. 4(b). The projected likelihood R_0 values using bootstrap resampling method with 1000 samples is shown in Fig. 4 (c). It is observed that minimum, median, mean and maximum values with 95% CI as 1.061, 1.161, 1.162 and 1.281 respectively. The epidemic trajectory with different quantile is shown in Fig. 4(d) for the next 10 days. The minimum, median, mean and maximum incidence in the next 10 days with 95% CI is observed as 607, 806, 804.9 and 1028 respectively. For the forecasted period, actual cases reported are 1242. The % error is obtained as 35.1% for 'earlyR' epidemic model with median forecasted value. The %error of the specific interval incidence is evaluated as,

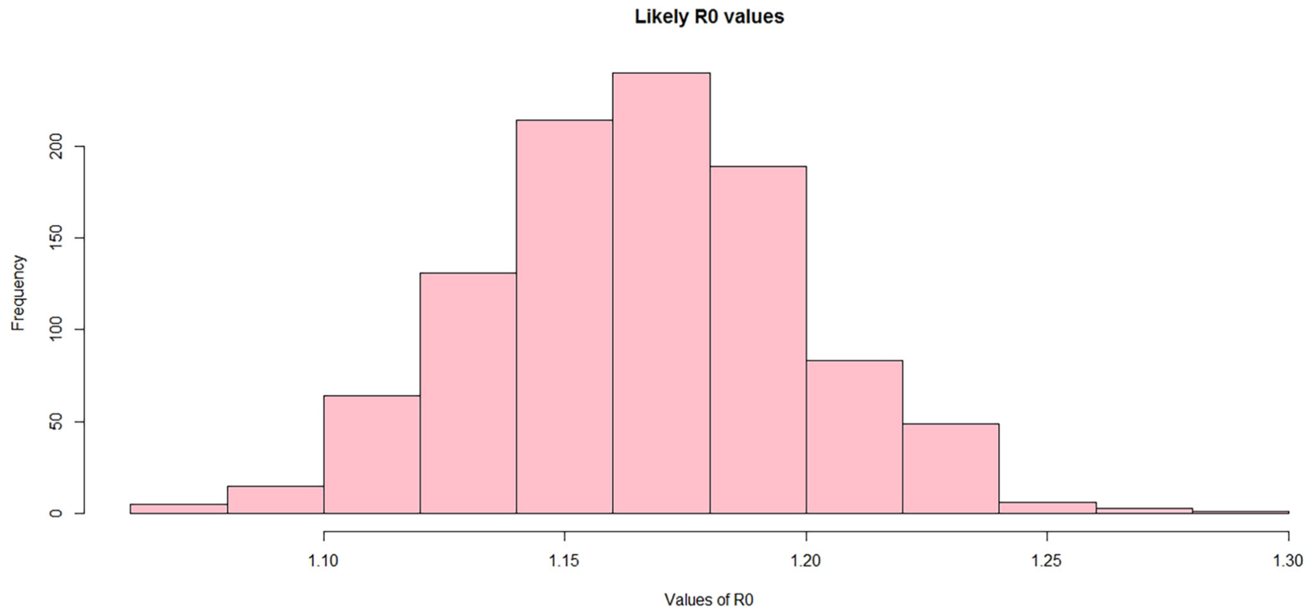

Fig. 4 (*continued*)

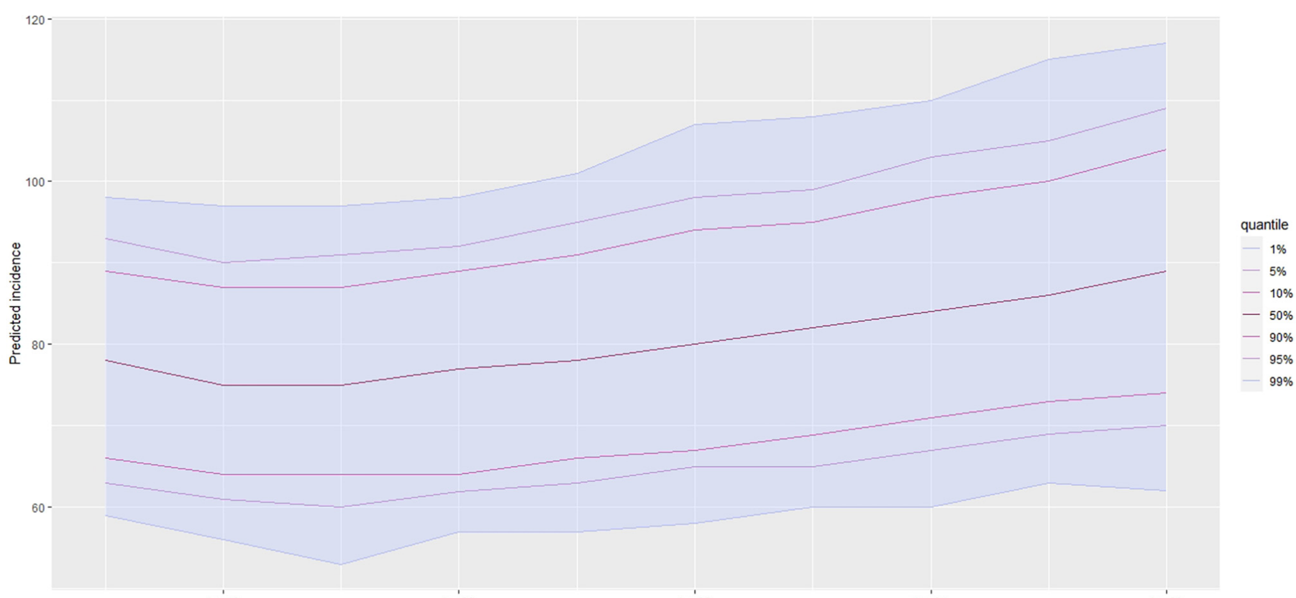

Fig. 4 (*continued*)

$$\%error = \frac{ActualCummulativeIncidence - PredictedCummulativeIncidence}{ActualCummulativeIncidence} * 100 \quad (5)$$

3.2. Evaluation of ARIMA model with real time data

The 'auto.arima' function in R is used and obtained model as ARIMA(0,1,1) for the India's COVID-19 data from 3rd May 2020 to 24th May 2020. The specified period data is available in Annexure I. Fig. 5(a) shows the actual and forecasted newly confirmed cases with 80% and 95% CI. The black line indicates the actual incidence. The blue line indicates the point of forecast. The blue and grey shades indicate the regions of 80% and 95% CI. The ACF and PACF plot of ARIMA evaluation model is shown in Fig. 5(b) and 5 (c) respectively. Table 1 shows the accuracy parameters of ARIMA evaluation model. It is observed that the model has been designed with 90.49%. It is estimated from (100-MAPE) value. Table 2 shows the comparison of actual reported cases with point of forecast value. It is observed the deviation ranging from −9.78491% to 11.61085%.

It is observed that compared to 'earlyR' epidemic model ARIMA model offers better accuracy with real incidence. The future pre-

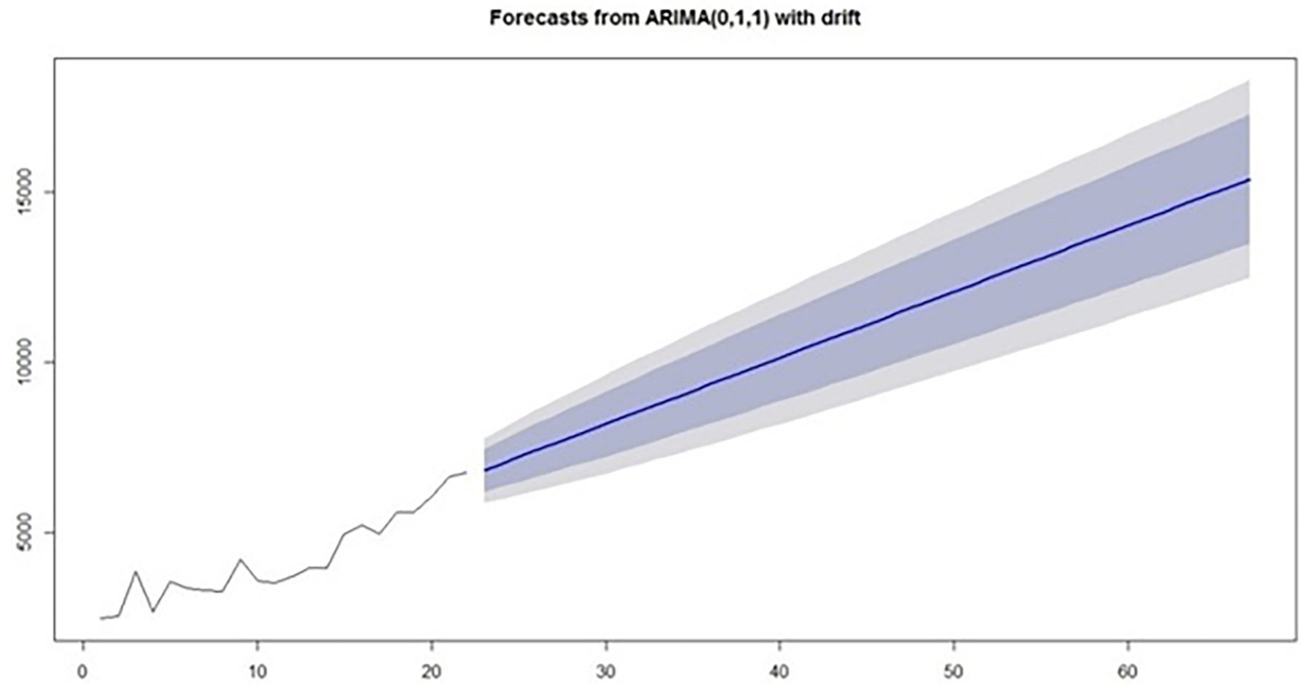

Fig. 5. (a). Actual and forecasted newly confirmed cases of evaluation model (b). ACF plot of evaluation model COVID-19 dataset (c). PACF plot of evaluation model COVID-19 dataset.

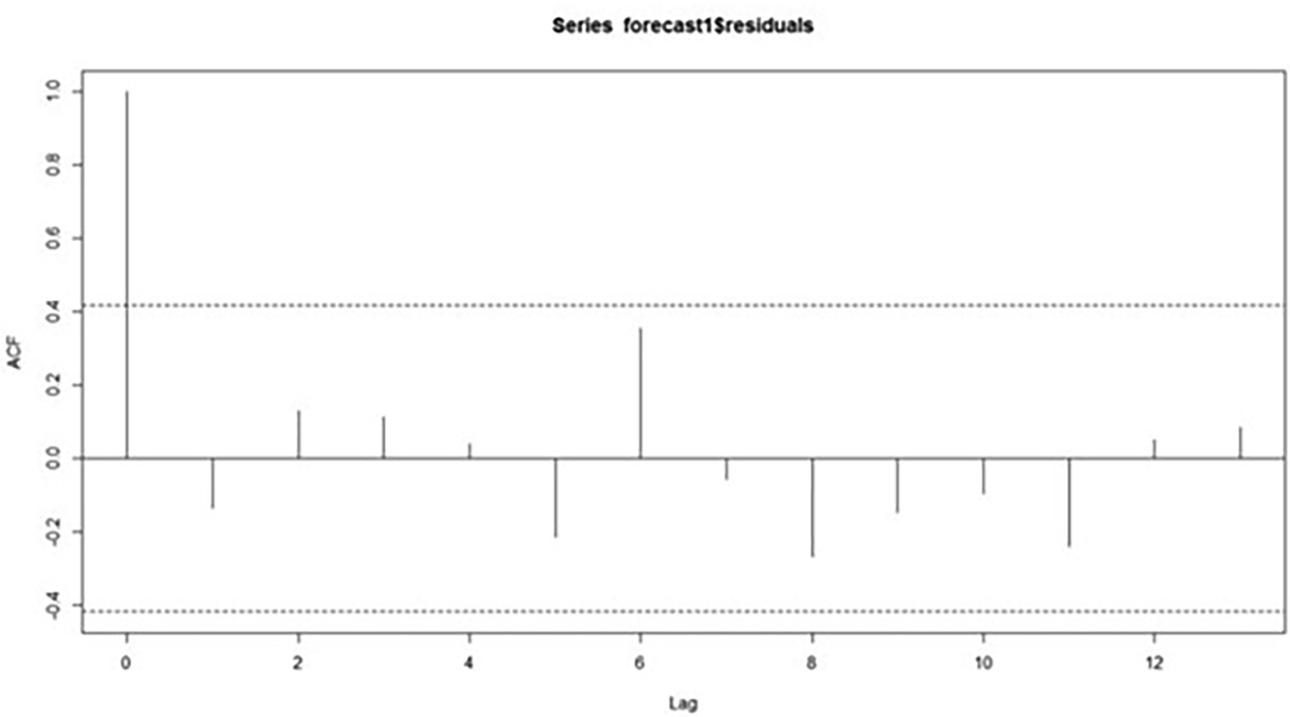

Fig. 5 (*continued*)

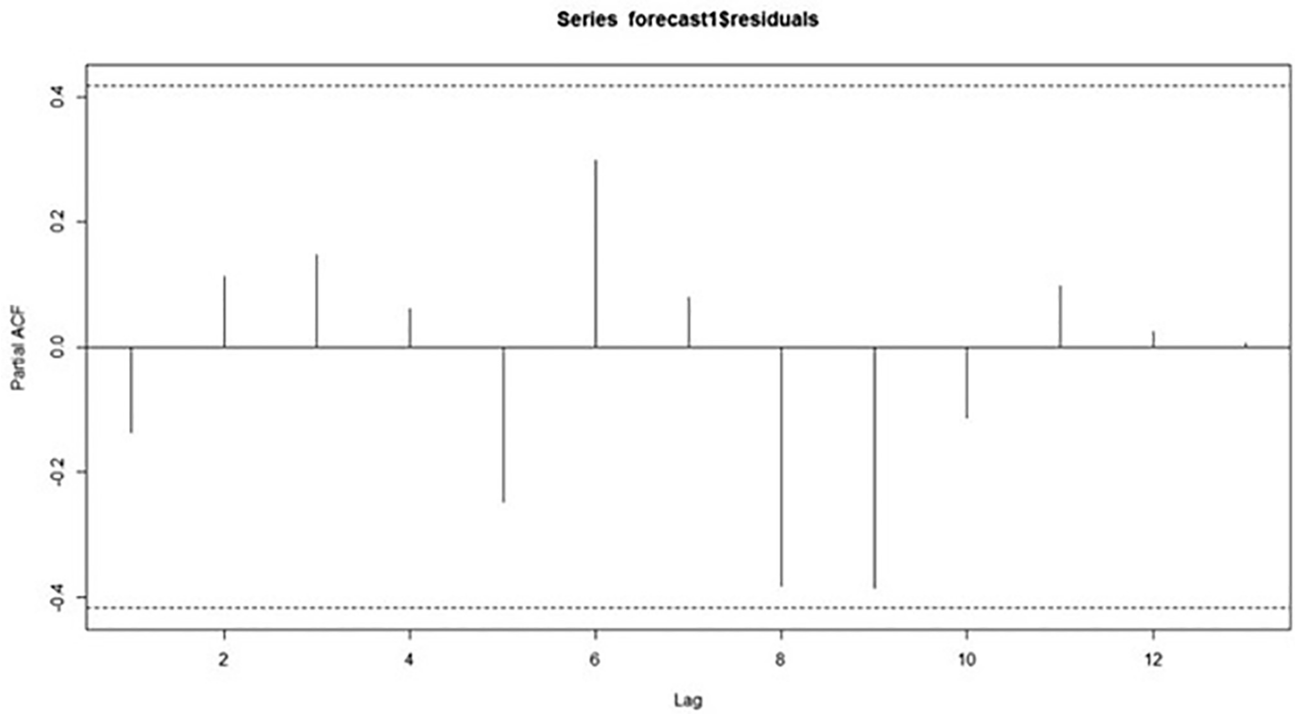

Fig. 5 (*continued*)

Table 1
Model accuracy parameters.

Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ARIMA (0,1,1) Training set	5.891605	449.6704	375.4912	−1.54434	9.514808	0.8606544	−0.1357024

dicted daily incidence purely depends on the reproduction number. The strict measures on infection management and control are required to bring down the infective cases in future failing may reach the maximum value. The WHO health advice should strictly be followed such as avoid touching eyes, nose and mouth, practicing the respiratory hygiene (wearing mask), seeking early medical care after feeling the symptoms of COVID-19, washing hands regularly and thoroughly with an alcohol-based hand sanitizers or wash them with soap and water and maintain social distance at least 1 m (3 feet) between them with the people those who are coughing / sneezing [13]. In addition, the quarantining the infected people with rapid diagnostics may also help to control the COVID-19 epidemic.

Table 2
Comparison Actual reported cases with predicted cases.

Date	Forecasted Cases	Actual reported cases	% Error
25-May	6838	6977	−2.03276
26-May	7032	6535	7.067691
27-May	7226	6387	11.61085
28-May	7420	6566	11.50943
29-May	7614	7466	1.943788
30-May	7809	7964	−1.98489
31-May	8003	8380	−4.71073
1-Jun	8197	8392	−2.37892
2-Jun	8391	8171	2.621857
3-Jun	8585	8909	−3.77402
4-Jun	8779	9304	−5.98018
5-Jun	8973	9851	−9.78491
6-Jun	9167	9887	−7.85426
7-Jun	9362	9971	−6.50502
8-Jun	9556	9983	−4.4684
9-Jun	9750	9987	−2.43077
10-Jun	9944	9985	−0.41231
11-Jun	10,138	9996	1.400671
12-Jun	10,332	10,956	−6.03949
13-Jun	10,526	11,458	−8.85427
14-Jun	10,720	11,929	−11.278
15-Jun	10,914	11,502	−5.38758
16-Jun	11,109	10,667	3.978756

4. Conclusion

The accuracy of the predicted figures depends on identified infective in the prescribed interval. The large number unidentified infective in the available daily incidence of data may bring uncertainties in the developed forecasting model. The strict measures will help the health policy makers to control the spread of COVID-19 epidemic. With the comparison of 'earlyR' epidemic model with ARIMA model, it is found that ARIMA model offers better accuracy. Our study will help the corresponding policy makers to take the precautionary measure using the suitable prediction model.

5. Data availability statement

The data used are publically available in the Indian government official website (https://www.mohfw.gov.in/). The extracted input data used for prediction is available in Annexure I.

Funding information

The research work is not funded by any agency / Institution

CRediT authorship contribution statement

Karthick K: Conceptualization, Data curation, Methodology. **Ebrahem A. Algehyne:** Formal analysis, Investigation. **Kavaskar Sekar:** Validation, Visualization, Writing - original draft, Writing -review & editing.

Declaration of 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.

References

[1] Alsharif, M.H.; Younes, M.K.; Kim, J. Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea. Symmetry 2019, 11, 240
[2] Anne Cori, Neil M. Ferguson, Christophe Fraser, Simon Cauchemez, A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, Am. J. Epidemiol. 178 (9) (2013) 1505–1512, https://doi.org/10.1093/aje/kwt133.
[3] COVID-19 Case history, 2020 link: http://covidindiaupdates.in/story_100cases.php.
[4] COVID-19 WHO report, 2020 link: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200618-covid-19-sitrep-150.pdf?sfvrsn=aa9fe9cf_2.
[5] Data Link, 2020: https://www.mohfw.gov.in/.
[6] Giraldo, J. Ospina, D. Hincapié Palacio, Deterministic SIR (Susceptible-Infected-Removed) models applied to varicella outbreaks, Epidemiol. Infection 136 (5) (2008) 679–687, https://doi.org/10.1017/S0950268807009260.
[7] Hiroshi Nishiura, Natalie M. Linton, Andrei R. Akhmetzhanov, Serial interval of novel coronavirus (COVID-19) infections, Int. J. Infect. Dis. 93 (2020) 284–286, https://doi.org/10.1016/j.ijid.2020.02.060.
[8] M. Kenji, C. Gerardo, Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020, Infect. Dis. Modell. 5 (2020) 264–270, https://doi.org/10.1016/j.idm.2020.02.003.
[9] Q. Li et al., Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia, New Engl. J. Med. 382 (13) (2020) 1199–1207, https://doi.org/10.1056/NEJMoa2001316.
[10] J. Pang, M.X. Wang, I.Y.H. Ang, S.H.X. Tan, R.F. Lewis, J.I. Chen, R.A. Gutierrez, S.X.W. Gwee, P.E.Y. Chua, Q. Yang, X.Y. Ng, R.K. Yap, H.Y. Tan, Y.Y. Teo, C.C. Tan, A. R. Cook, J.C. Yap, L.Y. Hsu, Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV): A Systematic Review, J Clin Med. 9 (3) (2020), https://doi.org/10.3390/jcm9030623, PMID: 32110875 PMCID: PMC7141113.
[11] Shunqing Xu, Yuanyuan Li, Beware of the second wave of COVID-19, The Lancet 395 (10233) (2020) 1321–1322, https://doi.org/10.1016/S0140-6736(20)30845-X.
[12] Thibaut Jombart, Pierre Nouvellet, Sangeeta Bhatia, Zhian N. Kamvarprojections: Project future case incidence (2018) Available from: https://CRAN.R-project.org/package=projections. [Accessed 27 August 2018]
[13] WHO Advice for Public, April, 2020 link: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public
[14] L. Zhong et al., Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model, IEEE Access 8 (2020) 51761–51769, https://doi.org/10.1109/ACCESS.2020.2979599.