Forecasting the COVID-19 Outbreak: An Application of Arima and Fuzzy Time Series Models

Prashant Verma
State Bank of India
Mukti Khetan (mukti.khetan11@gmail.com)
Indian Institute of Technology Bombay
Shikha Dwivedi
University of the Chinese Academy of Sciences
Shweta Dixit
GD Goenka University

Research

Keywords: Coronavirus, COVID-19, ARIMA, FUZZY Time Series, Forecasting

Posted Date: June 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36585/v1

FORECASTING THE COVID-19 OUTBREAK: AN APPLICATION OF ARIMA AND FUZZY TIME SERIES MODELS

Prashant Verma[1] Mukti Khetan[2] Shikha Dwivedi[3] and Shweta Dixit[4]

[1]Analytics Department Global IT Centre, SBI, Navi Mumbai, India
Email: prashantvermag@gmail.com
[2*]Department of Mathematics, Indian Institute of Technology Bombay, Mumbai Maharashtra, India
Email: mukti.khetan11@gmail.com
[3]Department of Physics, University of Chinese Academy of Sciences, Beijing, China
Email: shikhadwij@gmail.com
[4]School of Management, GD Goenka University, Gurugram, Haryana, India
Email: sdshwetadixit@gmail.com

Abstract

Purpose: The whole world is surfaced with an inordinate challenge of mankind due to COVID-19, caused by 2019 novel coronavirus (SARS-CoV-2). After taking hundreds of thousands of lives, millions of people are still in the substantial grasp of this virus. This virus is highly contagious with reproduction number R_0, as high as 6.5 worldwide and between 1.5 to 2.6 in India. So, the number of total infections and the number of deaths will get a day-to-day hike until the curve flattens. Under the current circumstances, it becomes inevitable to develop a model, which can anticipate future morbidities, recoveries, and deaths.
Methods: We have developed some models based on ARIMA and FUZZY time series methodology for the forecasting of COVID-19 infections, mortalities and recoveries in India and Maharashtra explicitly, which is the most affected state in India, following the COVID-19 statistics till "Lockdown 3.0" (17th May 2020).
Results: Both models suggest that there will be an exponential uplift in COVID-19 cases in the near future. We have forecasted the COVID-19 data set for next seven days. The forecasted values are in good agreement with real ones for all six COVID-19 scenarios for Maharashtra and India as a whole as well.
Conclusion: The forecasts for the ARIMA and FUZZY time series models will be useful for the policymakers of the health care systems so that the system and the medical personnel can be prepared to combat the pandemic.

Keywords: Coronavirus; COVID-19; ARIMA; FUZZY Time Series; Forecasting.

1. Background

No one could ever presume the prevailing scenario that a minuscule RNA structure can ingest millions of lives and still starved to have many more millions, leaving no stone unturned. The world, utterly equipped with leading-edge technology, is experiencing a disaster in saving lives as well as in screening the economy. Thus, despite warnings of some researchers, no country

was ever prepared for this confrontation with the imperceptible adversary of humanity. The ongoing pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), also acknowledged as 2019 novel coronavirus (2019-nCov) or COVID-19 but publicly known as just "coronavirus", the name derived from the virus structure, where spike protein looks like a shape of a crown [1]. The full spectrum of COVID-19 ranges from the common cold, fever, mild respiratory complications to severe progressive pneumonia, multi-organ failure, and death. This virus is devilishly transmissible with the reproduction number (R_0) ranging from 1.5 to 6.5 globally, i.e., each infected individual can infect at least 1.5 other individuals [2,3]. Thus, each infection causes furthermore and the outbreak will continue to grow. The estimated number of R_0 in India was 2.56 prior to the "Lockdown 2.0" (30th April 2020) [4] and it appears that Maharashtra has an elevated R_0, as there, growth in the number of COVID-19 victims is intense, particularly, in two of its robust cities, Mumbai and Pune.

Even now, it is un-conceivable that this virus has spread its deadly wings around the whole world and gripped all 218 countries into its lethal clutches in such a short span of time. Thus far, over five million human beings have been infected, and more than 0.3 million lives have been lost worldwide [5]. In this era, many economic, medical, political, and cultural activities, as well as the tourism enterprises of various countries, are interconnected with each other, which led the virus to spread so fast. India also made no exemption and got its first COVID-19 case as early as on the 30th of January 2020 in Kerala. At the initiation stage, the dissemination was slow as only three COVID-19 cases were identified for more than a month but after that, morbidities got an exponential surge and then took no halt since then. Until the 3rd national lockdown i.e. "Lockdown 3.0" (17th May 2020), almost 0.1 million confirmed cases had been reported in India, of which over 3000 succumbed to the infection [6]. With these statistics, India is confronting a grave threat of COVID-19 disease and set foot into the list of the top 10 most infected countries of the world. Maharashtra is the worst affected state of India, which alone accounts for one-third of the country's total cases.

The ultimate solution to this crisis is the successful development of the vaccine. However, as of now, the only alternative is to contain the disease, preventing its further diffusion in society. In this connection, we need some temporary makeshifts to properly handle the morbidities with currently available options as the startling increase in COVID-19 cases is putting tremendous pressure on the medical services, leading to a shortage of intensive care resources throughout the country. Under such background, the utmost requirement is to establish a realistic and computationally effective methodology that can successfully forecast the potential morbidities, and thus, it can assist in decision making and logistical planning in healthcare systems for the following imminent challenges. These models may curtail the anticipated chaos for health care workers. The statistical prediction models are always conducive to prediction and in this way, it may restrain the comprehensive plight in the pandemic.

In the present work, we have used the Auto-Regressive Integrated Moving Average (ARIMA) model [7] and the fuzzy time series (FTS) model [8] for determining the forthcoming number of COVID-19 cases of India as a whole, along with India's worst-hit state, Maharashtra. The rationale behind the choice of the ARIMA model is the following:

i. This model gives remarkably satisfying results in the prediction of natural adversities, when compared to other prediction models, such as wavelet neural network (WNN) model and support vector machine (SVM) model [9].

ii. ARIMA model has been used earlier in a similar crisis like situation, as during the SARS outbreak, Earnest et al. [10] exercised this model to make a real-time prediction on the number of beds occupied in Tan Tock Seng Hospital, Singapore. Also, Anwar et al. [11] used the ARIMA model to develop a predictive tool for malaria patterns in Afghanistan. Recently, Benvenuto et al. [12] employed this model to anticipate the forthcoming COVID-19 morbidities globally.

Although limited research has been done on FTS methodology in health and epidemiology fields, yet it can be an alternative way to forecast the COVID-19 sepsis, as for some study, the FTS model proved itself more precise than ARIMA model. For instance, Wang [13] shows that the FTS model can be utilized to predict export values (from Taiwan) accurately, outperforming the ARIMA model. Similar conclusions have been drowned by Zhang et al. [14] in the analysis of infectious disease surveillance and by Tricahya and Rustam [15] in forecasting the number of pneumonia patients in Jakarta.

We have implemented these two methodologies and COVID-19 stats from the Indian Government database [6] (up to "Lockdown 3.0") for calculations and forecasted the COVID-19 infections, mortalities, and recoveries. The paper is organized as follows: In Section 2, we recount the details of ARIMA and FTS models used in this work. It is followed by the results of modeling and a detailed analysis of the outcome, in Section 3. Finally, conclusions are drawn in Section 4.

2. Material and methods

This study utilizes the COVID 19 data obtained from the 'Ministry of Health and Family Welfare, Government of India' concerning new cases/infections, mortalities, and recoveries[6]. In order to forecast the COVID 19 scenarios in India, classical time series (ARIMA) and fuzzy time series have been utilized. The core difference between the classical time series and the fuzzy time series is that the values of the former are real numbers while the values of the latter are fuzzy sets.

2.1. ARIMA model

ARIMA models (Box–Jenkins, [7]) are the extensively used technique for time series forecasting, characterizing the autocorrelations in the data. The algorithm of this model is as follows:

2.1.1 Identification of stationary time series:

a) The first step in developing the Box–Jenkins model is to determine whether the data is stationary or not. If nonstationarity exists, it can be modeled by differencing it to an appropriate level of difference. The Augmented Dickey-Fuller (ADF) test is performed to test the stationarity of the data. If the p-value is greater than 0.05, we reject the null hypothesis (H_0: unit root is present, i.e., non-stationary) otherwise it will be accepted at a 5% level of significance.

b) Identification of parameters: The next step is to identify the parameters p and q of the ARMA (p, q) model by using two diagnostic plots, autocorrelation function (ACF) and partial autocorrelation function (PACF). For a given, non-stationary time series X_t, the ARIMA (p, d, q) model can be written as:

$$\phi(B)(1-B)^d X_t = \psi(B)\varepsilon_t$$

Where,

d: Number of times, need to differentiate X_t, to make the data stationary.

B: Backward shift operator.

Hence,

$\phi(B): 1-\phi_1 B-\phi_2 B^2-...-\phi_p B^p$

$\psi(B): 1+\psi_1 B+\psi_2 B^2+...+\psi_p B^p$ are characteristic polynomials in B and ε_t is the white noise, where, $\varepsilon_t \sim N(0,1)$

2.1.2 Parameter estimation: *R 3.6.1* software has been used to find coefficients that are best fitted for the selected ARIMA model. The selection of the ARIMA model is done by the minimum Akaike information criterion (AIC), [16]. It is written as

$$AIC = -2\log(L) + 2(p + q + k),$$

Where L is the likelihood of data, p is the order of autoregressive part, q is the order of moving average part, and k represents the intercept of the ARIMA model.

2.1.3 Diagnostic check: This process involves testing of the assumptions of models: whether the parameters achieve statistical significance or multicollinearity and whether the residual term is white noise or not. Ljung- Box tests are performed to test the adequacy, and ACF of the residuals are plotted. The aforementioned steps are reiterated until the required adequacy is achieved.

2.2 Fuzzy time series model: The concept of the fuzzy time series was first developed by Song and Chisson [8]. The main advantage of the FTS technique is that there are no assumptions considered for the data set. Here, we briefly reviewed some concepts of FTS from [8, 17].

Definition 1 (Fuzzy Set) [18]

Let U be the universe of discourse, where $U=\{u_1,u_2,...u_n\}$. A fuzzy set A_i of U is defined by $A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + ... + f_{A_i}(u_n)/u_n$, where f_{A_i} is the membership function of fuzzy

set A_i, $f_{A_i}: U \rightarrow [0,1]$. u_k is the element of fuzzy set A_i, and $f_{A_i}(u_k)$ is the degree of belongingness of u_k to A_i. $f_{A_i}(u_k) \in [0,1]$ where $1 \le k \le n$. The definition of FTS is briefly reviewed as follows:

Definition 2 (FTS) [8, 17]
Let $Y(t)(t = ..., 0, 1, 2...)$, is a subset of R. Let $Y(t)$ be the universe of discourse defined by fuzzy set $f_i(t)$. If $F(t)$ collection of $f_i(t)$ $(i = 1, 2, ...)$, $F(t)$ is defined as a fuzzy time series on $Y(t)$ (t = ..., 0, 1, 2, ...).

Definition 3 (Fuzzy logical relationship) [8, 17, 19]
If there exists a fuzzy relationship $R(t-1,t)$, such that $F(t) = F(t-1)$ x $R(t-1,t)$ where x represents an operator, then $F(t)$ is said to be caused by $F(t-1)$. When $F(t-1) = A_i$ and $F(t) = A_j$, the relationship between $F(t-1)$ and $F(t)$ is called a fuzzy logical relationship (FLR) and denoted by $A_i \rightarrow A_j$.

Definition 4 (fuzzy logical relationship group) [8, 17, 19]
Fuzzy logical relationships with the same fuzzy set on the left-hand side can be further grouped into fuzzy logical relationship groups. Suppose there are fuzzy logical relationships such that

$$A_i \rightarrow A_{j1},$$
$$A_i \rightarrow A_{j2},$$
$$A_i \rightarrow A_{jm},$$

They can be grouped into a fuzzy logical relationship group (FLRG) $A_i \rightarrow A_{j1}, A_{j2}, ...A_{jm}$

We have estimated the forecast values of COVID-19 data sets by following Chen's model [19], but the length of the interval is defined as the algorithm average based length given by Hurang [20].

The algorithm of FTS model is presented as follows:

Step 1 Define the universe: The universe of discourse U[L_c - D_1, U_c + D_2], where L_c and U_c are the lower and Upper limit of the data respectively and D_1, and D_2 are the constants which we get using 10% of the L_c and U_c. So, we have extrapolated the upper and lower bounds by 10% as a security margin.

Step 2 Divide the universe of discourse into intervals: The Universe of discourse is partitioned into *'l'* length of the interval, where *'l'* is based on algorithm 'average based length' by Hurang (2001). Using this length *'l'*, we divide the universe of discourse into intervals.

Step 3 Fuzzify the time series data set: Achieve $u_1, u_2, ...u_n,$ and define fuzzy sets A_i; $A_1, A_2, ...A_n$ and fuzzify the data.

Step 4 Identify the Fuzzy Logical Relationship and Fuzzy Logical Relationship Group: Identification of the FLRs and FLRGs based on Definition 3 and 4 respectively.

Step 5 Defuzzify and compute the forecasted values: We have used the algorithm of Chen's model [19] for forecasting and result are presented in Table 2.

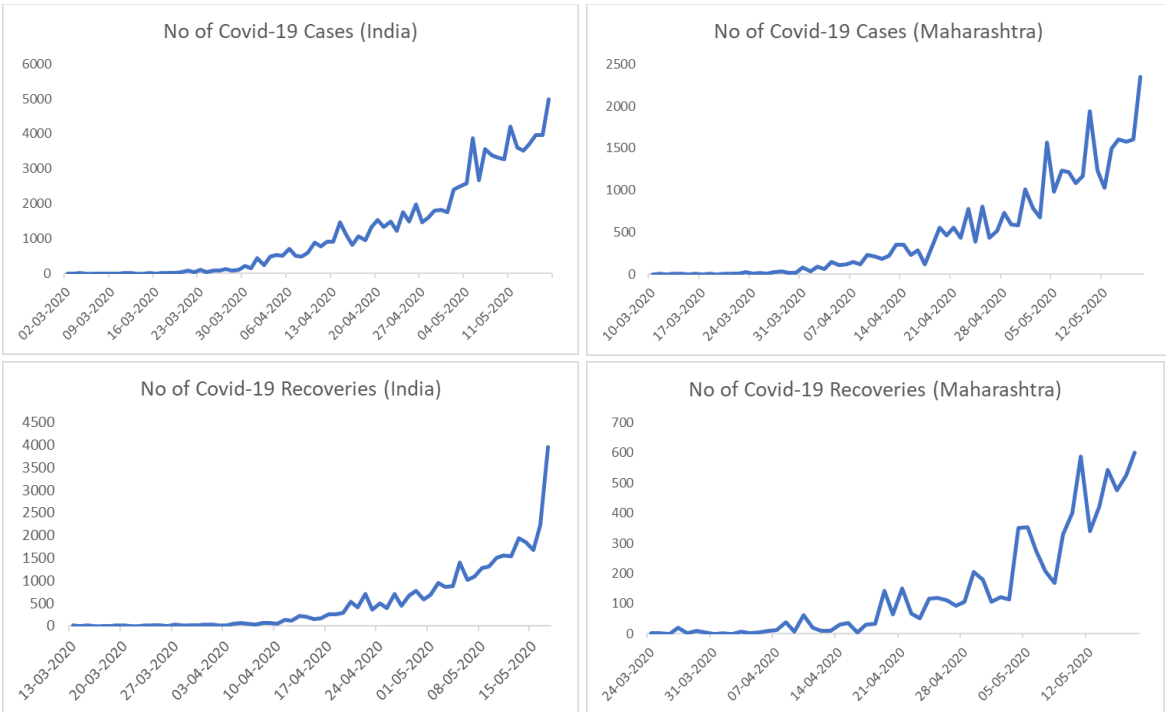
Step 6 Validation: Examine the performance analysis parameter using Root Mean Square Error [RMSE]. The RMSE can be defined as

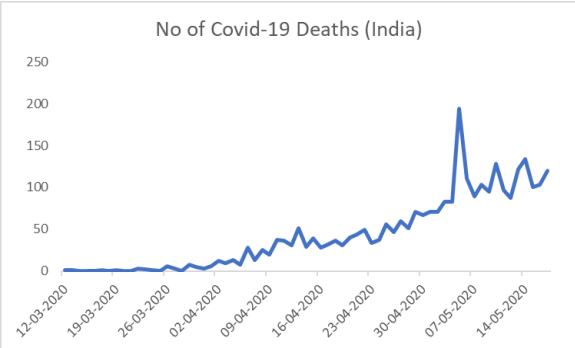
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Forcast_i - Actual_i)^2}{n}}$$

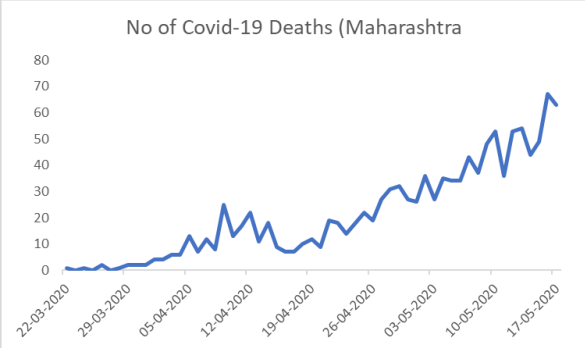
3. Results and Discussions

The data sets utilized in the forecasting, are collected from various Indian databases [6], which contain total COVID-19 morbidities, recoveries, and deaths, reported in India and also contain data subsets for the state of Maharashtra throughout the phase between 2nd March 2020 to 17th May 2020. In Figure 1, we have plotted six distinct datasets for daily cases of COVID-19 infections, recoveries, and deaths recorded in India and Maharashtra, which shows an exponential growth for all the six data sets. Thus we can say that all the series are *non-stationary*. We have also conducted the Augmented Dickey-Fuller / Unit root test to prove that these six data sets are not stationary.

Figure 1: Graphical Representation of COVID-19 statistics in India and Maharashtra







3.1 Unit root test/ADF Test

We have performed an ADF test to check the stationarity of data and results of the test are summarized in Table 1 showing p-values corresponding to different datasets. According to the ADF test, p-value > 0.05, leads to a non-stationary model, and as it can be seen from the table, that the p-value corresponding to each data sample, is greater than 0.05, implying that all the series are non-stationary, which validates the conclusion from previous graphs.

Table 1: Unit Root Test of the six data sets of COVID-19

Data set	Test statistic value	P-value
India		
No of cases	-0.31422	0.9876
No of recovery	3.09040	0.9900
No of deaths	-2.06880	0.5472
Maharashtra		
No of cases	0.024982	0.9900
No of recovery	-0.69697	0.9652
No of deaths	-0.59160	0.9744

3.2 Autocorrelation function and Partial Autocorrelation function: ACF and PACF of the data have been used for assessing the combination of ARIMA models. We have used this method to find the values of parameters p and q.

3.3. ARIMA and FTS model fitting: The best fitted ARIMA and FTS models for COVID-19 datasets of India and Maharashtra along with the estimated parameters are summarized in Table 2. The diagnostic measures AIC and RMSE are also mentioned. From this table, it is clear that among the daily cases of COVID-19 infections and recoveries in India particularly in Maharashtra, the ARIMA model appears to be a more suitable model as it gives lower RMSE value compared to FTS model. Similarily, ARIMA predicts better for the case of COVID-19

deaths in India. However, in the case of COVID-19 deaths in Maharashtra, the FTS model is the best-fitted model for forecasting because of its lower RMSE value.

Table 2: Best fitted model of COVID-19 data sets with parameter estimation, diagnostic measure, and RMSE values

Dataset	Model	Coefficients					AIC	RMSE (ARIMA)	RMSE (FTS)
No of Covid-19 Cases (India)	(2,1,2)		AR1	AR2	MA1	MA2	1082	277	298.53
			-0.7560	-0.7017	0.2891	0.7103			
		S.E.	0.1535	0.1634	0.1654	0.1427			
No of Covid-19 Recoveries (India)	(1,2,1)		AR1	MA1			900	255.71	247.58
			-0.0428	-0.8386					
		S.E.	0.2048	0.1228					
No of Covid-19 Deaths (India)	(2,1,2)		AR1	AR2	MA1	MA2	574	17.77	16.67
			0.9243	-0.1613	-1.5693	0.7018			
		S.E.	0.2335	0.1829	0.1977	0.1570			
No of Covid-19 Cases (Maharashtra)	(1,1,2)		AR1	MA1	MA2		921	196.22	198.67
			0.4967	-1.3323	0.7081				
		S.E.	0.1910	0.1441	0.1313				
No of Covid-19 Recoveries (Maharashtra)	(3,1,2)		AR1	AR2	AR3	MA1	MA2	610	58.30	64.91
			1.0087	-0.6105	-0.0518	-1.5802	1.000			
		S.E.	0.1411	0.1706	0.1432	0.0879	0.0995			
No of Covid-19 Deaths (Maharashtra)	(2,1,2)		AR1	AR2	MA1	MA2	359	5.15	5.16
			0.1301	-0.5676	-0.7439	0.9999			
		S.E.	0.1180	0.1179	0.0785	0.1819			

3.4 Diagnostic Check

3.4.1 Ljung Box test for residuals: The Box-Ljung test [7] is a diagnostic tool used to test the lack of fit of a time series model. In the following Table 3, we have shown the values of the Ljung-Box test statistic along with its corresponding p-value for all the six data sets.

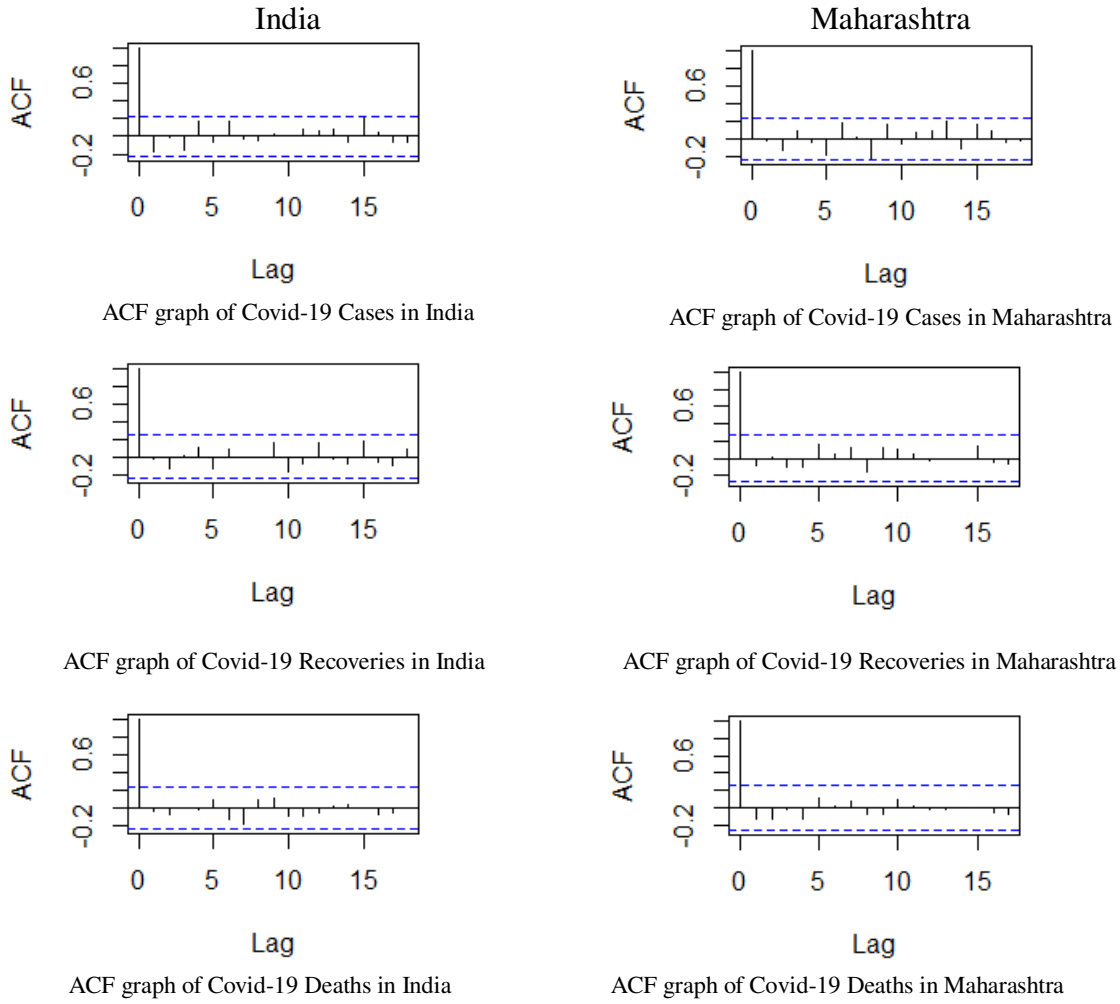
Table 3: Ljung-Box Test for COVID-19 data

Dataset	Ljung Box test statistic	p-value
No of Covid-19 Cases (India)	17.888	0.5299
No of Covid-19 Recoveries (India)	16.590	0.6177
No of Covid-19 Deaths (India)	9.0866	0.9720
No of Covid-19 Cases (Maharashtra)	22.232	0.2729
No of Covid-19 Recoveries (Maharashtra)	-0.69697	0.9652
No of Covid-19 Deaths (Maharashtra)	11.736	0.8966

Table 3 provides relevant validation ($p > 0.5$) in favor of the Null Hypothesis at 5% level of significance among all the six cases establishing the suitability of the models.

3.4.2 Residual ACF: As we know, the primary measure of goodness of fit is that the errors of the fitted models should be uncorrelated. This is the required condition which ensures that the model fitting is appropriate. We have plotted the residual ACFs in Figure 2, where black spikes in between the horizontal dotted blue lines (control band) are the residual ACFs. If the spikes of residual ACF are on both sides of the horizontal axis in a random pattern and gradually decreases to zero, we may call it an ideal situation. Eventually, as the lag increases, spikes fall inside the control band. Hence, it is clear from Figure 2 that all ARIMA models are fitted appropriately.

Figure 2: Graphical representation of residual ACF of COVID-19 data samples



ACF graph of Covid-19 Cases in India

ACF graph of Covid-19 Cases in Maharashtra

ACF graph of Covid-19 Recoveries in India

ACF graph of Covid-19 Recoveries in Maharashtra

ACF graph of Covid-19 Deaths in India

ACF graph of Covid-19 Deaths in Maharashtra

3.5 Forecasting

The forecasted values of COVID-19 infections, recoveries, and deaths in India and Maharashtra respectively, have been compiled in Table 4. The prediction is based on the best-fitted model with minimum RMSE. Thus, in cases of daily COVID-19 infections and recoveries in India and in Maharashtra along with the COVID-19 deaths in India, we have shown the prediction of ARIMA models in terms of lower (Lo) and upper (Hi) predictive intervals and FTS forecast are shown for the cases of COVID-19 deaths in Maharashtra.

Table 4: Forecasted values of COVID-19 pandemic based on ARIMA and FTS model

No. of Cases (India)					
Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
18-05-2020	4574.929	4217.354	4932.504	4028.065	5121.793
19-05-2020	4616.351	4211.141	5021.561	3996.635	5236.066
20-05-2020	4874.169	4357.882	5390.457	4084.576	5663.763
21-05-2020	4650.194	4032.384	5268.004	3705.335	5595.053
22-05-2020	4638.619	3978.005	5299.234	3628.296	5648.942
23-05-2020	4804.524	4078.981	5530.067	3694.902	5914.146
24-05-2020	4687.221	3892.622	5481.819	3471.987	5902.455

No. of Recoveries (India)					
Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
18-05-2020	4269.975	3937.176	4602.773	3761.004	4778.946
19-05-2020	4644.256	4144.924	5143.588	3880.594	5407.918
20-05-2020	5015.956	4360.713	5671.198	4013.848	6018.063
21-05-2020	5387.766	4578.021	6197.511	4149.368	6626.164
22-05-2020	5759.572	4793.296	6725.847	4281.780	7237.363
23-05-2020	6131.377	5005.033	7257.722	4408.782	7853.973
24-05-2020	6503.183	5212.493	7793.874	4529.243	8477.124

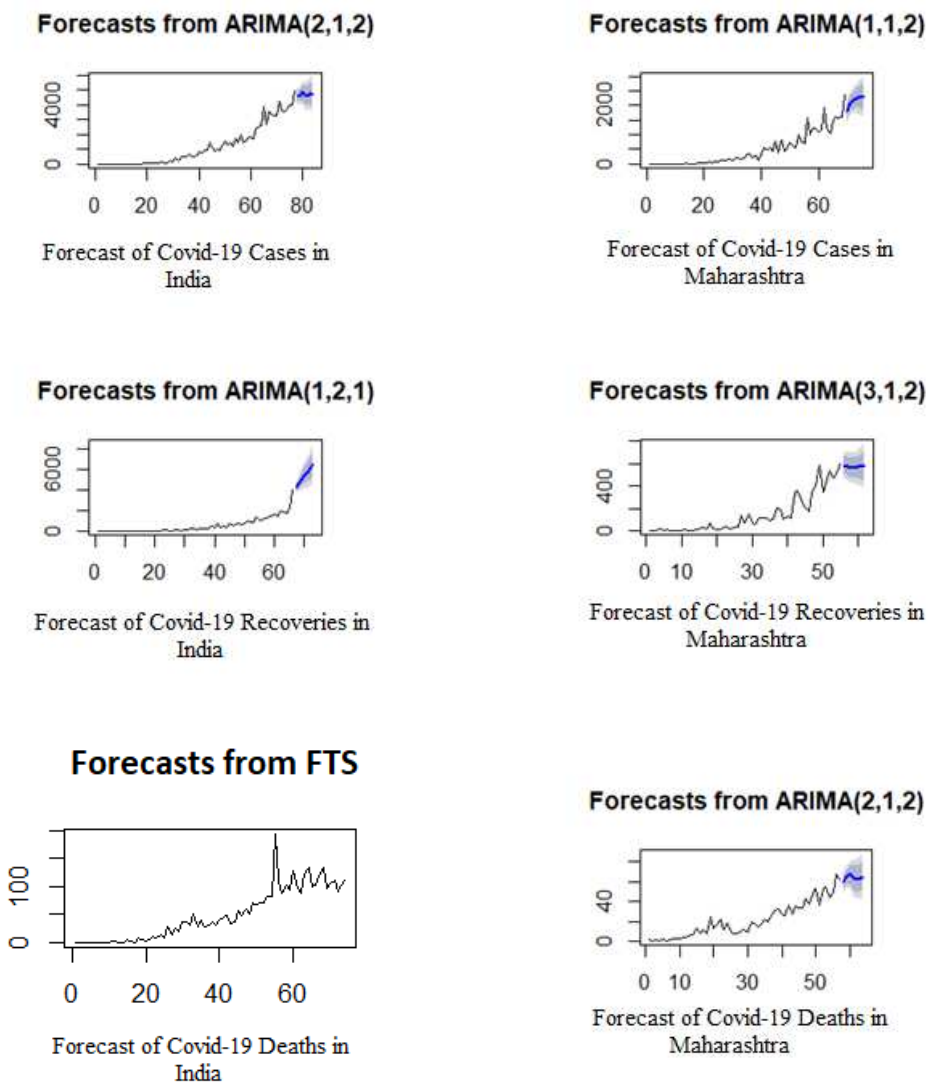
No. of Cases (Maharashtra)					
Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
18-05-2020	1783.557	1530.237	2036.877	1396.137	2170.977
19-05-2020	2047.780	1791.059	2304.501	1655.159	2440.401
20-05-2020	2179.017	1897.348	2460.685	1748.242	2609.792
21-05-2020	2244.201	1923.770	2564.631	1754.144	2734.257
22-05-2020	2276.577	1913.332	2639.821	1721.042	2832.111
23-05-2020	2292.657	1887.179	2698.136	1672.531	2912.783
24-05-2020	2300.644	1855.080	2746.209	1619.212	2982.077

No. of Recoveries (Maharashtra)					
Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
18-05-2020	589.8314	513.1324	666.5304	472.5304	707.1325
19-05-2020	577.4383	493.2116	661.6650	448.6246	706.2520
20-05-2020	567.2070	480.6949	653.7190	434.8983	699.5156
21-05-2020	564.9792	474.5108	655.4477	426.6197	703.3388
22-05-2020	569.6203	468.8292	670.4115	415.4735	723.7671
23-05-2020	576.1919	459.2873	693.0964	397.4018	754.9819
24-05-2020	580.1027	447.3557	712.8498	377.0836	783.1219

	No. of Deaths (Maharashtra)					No. of Deaths (India)
Date	Forecast (ARIMA)	Lo 80	Hi 80	Lo 95	Hi 95	Forecast (FTS)
18-05-2020	59.76160	52.98000	66.54319	49.39004	70.13315	134.137
19-05-2020	64.80111	57.50464	72.09757	53.64213	75.96009	97.5540
20-05-2020	67.29481	58.51826	76.07136	53.87223	80.71738	105.683
21-05-2020	64.75885	53.22811	76.28959	47.12410	82.39359	109.748
22-05-2020	63.01353	49.73721	76.28986	42.70914	83.31792	91.4570
23-05-2020	64.22586	50.04374	78.40798	42.53618	85.91554	103.651
24-05-2020	65.37420	50.23762	80.51079	42.22479	88.52362	109.748

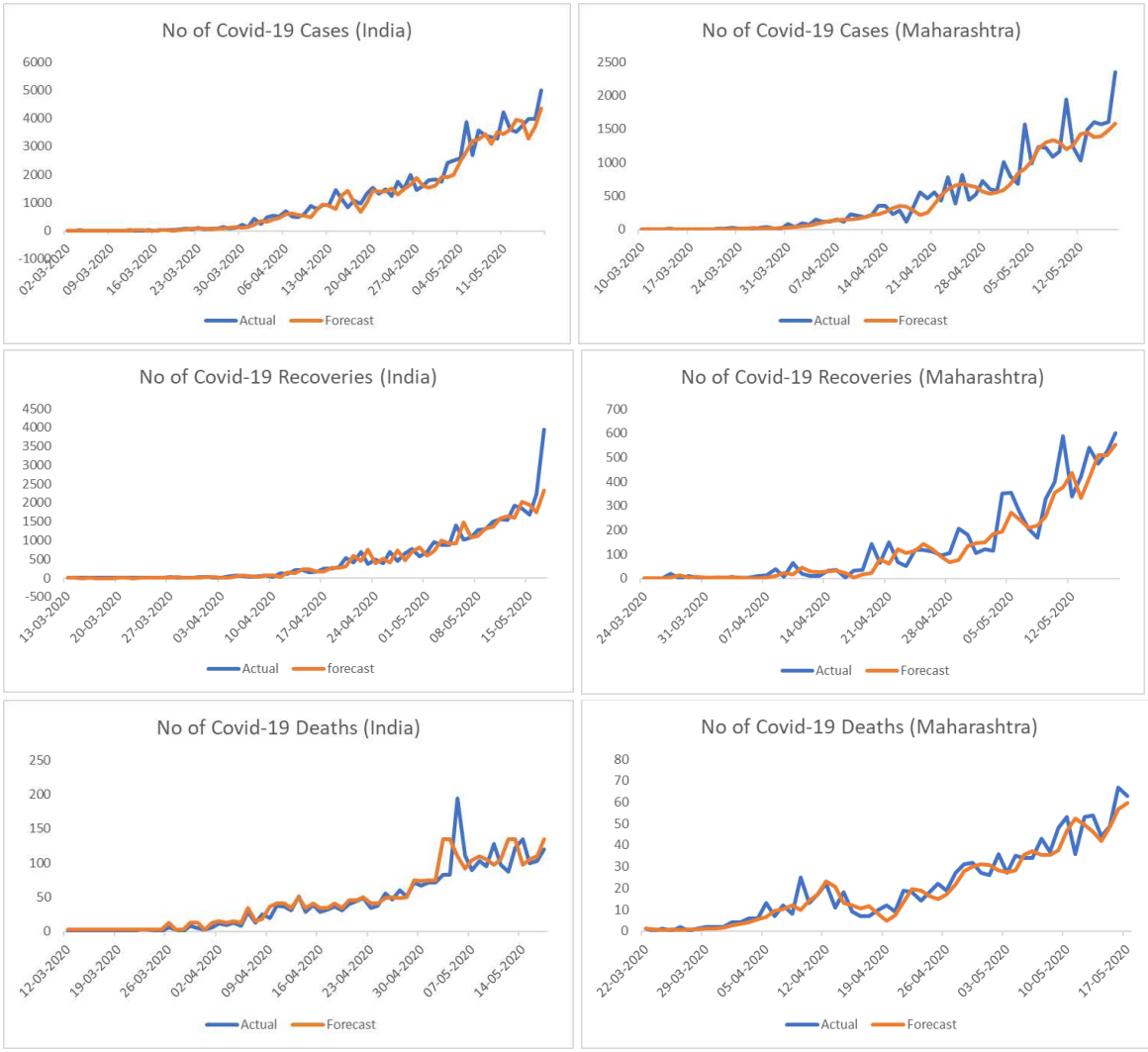
L_0, H_i : lower and upper predictive intervals with the risk error (size of the critical region): α = 0.2, 0.05

Figure 3: Graphical representation of Forecasted values of the COVID-19 pandemic with ARIMA and FTS Model



In Figure 3 we have plotted the forecasts of COVID-19 outbreak in India and Maharashtra resulted from ARIMA models (5 sets) and FTS model (1 set), using data as of Table 4. In this figure, the dark blue line shows the actual forecasted values whereas the ensemble of light blue lines gives the range of estimated values between the lower and upper predictive intervals.

Figure 4: Comparison between the actual and forecasted values of COVID-19 data sets



In Figure 4, we have plotted the actual values against the forecasted values of the parameter concerning COVID-19 pandemic data with ARIMA and FTS model. We have utilized *RStudio (Version 1.2.5)* for forcasting COVID-19 data through ARIMA model and *Python (Version 3.7)* through the Fuzzy Time Series model.

4 Conclusions

We have used univariate time series models (ARIMA and FTS) to estimate all aspects of the looming COVID-19 infections, recoveries, and deaths in India and its most vulnerable state of the ongoing pandemic, Maharashtra, based on contemporary statistics of gratis Indian data

banks. In this rationale, first, we have modeled three different COVID-19 non-stationary samples of infections, recoveries, and deaths in India and Maharashtra separately. Thereafter, we have fitted the model parameters in accordance with both ARIMA and FTS methodology. The calculated RMSE values of these models are utilized as conclusive tools to demonstrate befitting models for all six COVID-19 data sets concerning three parameters, incidence, mortality, and recovery. We have reported the estimated range of COVID-19 figures by ARIMA models, including randomized risk error factors (Figure 3). This report shows that hereafter the risk of contamination in India is not likely to decrease, particularly in Maharashtra, we may see an intensified scenario. Although the speculated overall recovery rate in India holds an optimistic trend for the upcoming days as the recovery rate is considerably higher than the number of new infections.

Nevertheless, this is not the case for Maharashtra, where the number of new infections will increase sharply and the recovery rate will not be so satisfactory in the following time. Death predictions are concluded by the ARIMA model explicitly in Maharashtra and by the FTS model in India as a whole (Figure 3). In both cases, the average mortality rate is persistent, under the global average.

We have compared the forecasted values with the actual statistics of COVID-19 till "Lockdown 3.0" to assess the potential of our models. Figure 4 depicts the forecasted values are in good agreement with real ones for all six COVID-19 scenarios for Maharashtra and India as a whole as well. However, the figure exhibits a few instances where we have observed some fluctuations between the observed and forecasted number of COVID-19 infections and recoveries on day to day basis. Our models may face this limitation in a few cases because of the following constraints:

1. ARIMA and FTS models reviewed the pandemic by considering the time variations of only one variable, while a number of effective tools that possess a role in the picture of correlation of COVID-19 data sets were extended in subsequent lockdowns such as testing capacity enhanced at least tenfold and still continue to increase.

2. Despite lockdown, millions of people have been migrated from their workplaces to hometowns, of which many found infected on arrival. So, these people were either the earlier undetected cases or they got infected during their journey.

3. Many infected individuals, particularly asymptomatic or mild ones are still not reported to the healthcare systems.

As per the findings of the current study, the risk of a boom in the infection rate will be higher in the coming days. Also, as of now, "Lockdown 4.0" is much more relaxed than the previous lockdowns, hereafter we will notice an uninterrupted growth in total COVID-19 cases of India and Maharashtra. This is an alarming state of affairs for the policymakers to practice better resources in prevention and control of the further transmission. Our models may assist them to estimate new infections and to plan the required strategy for supplying potential health facilities to all individuals and for balancing the burden of public health care systems as the majority of hospitals of India are established in urban areas only. This preparatory preparation will heal

the current adversity, protect human lives, and prevent heavy economic losses. Realizing the crucial importance of such modeling, we are planning to extend this work and use different multivariate models for forecasting, which will provide a further precise description of the current pandemic situation.

List of abbreviations: Not applicable.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: Data are available on www.mohfw.gov.in, www.covid19india.org

Competing interests: "The authors declare that they have no competing interests".

Funding: Not applicable.

Authors' contributions: Prashant Verma participated in the coding using R and python. Mukti khetan participated in the design of the study and performed the statistical analysis. Shikha Dwivedi reviewed the literature and drafted the manuscript. Shweta Dixit participated in the theoretical section of ARIMA and FUZZY TIME SERIES model. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

References

1. Tyrrell D.A. and Fielder M. (2002). *Cold Wars: The Fight Against the Common Cold*. Oxford UniversityPress. p. 96. ISBN 978-0-19-263285-2

2. James Holland Jones (2007). Califonia: Department of Anthropological Sciences.

3. Kevin Linka, Mathias Peirlinck, Ellen Kuhl, Reproduction number of COVID-19 and how it relates to public health measure, Department of Mechanical Engineering, Stanford University, Stanford, California, United States, medRxiv preprint DOI: https://doi.org/10.1101/2020.05.01.20088047.

4. Balram Rai, Anandi Shukla, Laxmi Kant Dwivedi, COVID-19 in India: Predictions, Reproduction Number and Public Health Preparedness, Department of Mathematical Demography and Statistics, International Institute for Population Sciences, Mumbai 400088 medRxiv preprint DOI: https://doi.org/10.1101/2020.04.09.20059261.

5. Website: www.worldometers.info/coronavirus/

6. Websites: www.mohfw.gov.in, www.covid19india.org

7. Box, G. E. P., and G. M. Jenkins (1976). Time Series Analysis: Forecasting and control. Rev. ed. San Francisco: Holden-Day

8. Song Q. and Chissom B.S. (1993). Forecasting enrollments with fuzzy time series—part I, Fuzzy Sets Syst 54(1):1–9.

9. Zhang Y., Yang H., Cui H., and Chen Q. (2019). Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the sanjiang Plain, China, Nat. Resour. Res., 29, 1447.

10. . Earnest, A., Chen, M. I., Ng, D., and Sin, L. Y. (2005). Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. BMC Health Services Research, 5(1), 36.

11. Anwar, M. Y., Lewnard, J. A., Parikh, S., and Pitzer, V. E. (2016). Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence, Malaria Journal, 15(1), 566.

12. Benvenuto D., Giovanetti M., Vassallo L., Angeletti S. and Ciccozzi M. (2020). Application of the ARIMA model on the COVID- 2019 epidemic dataset, Data in brief 29:105340.

13. Wang C.C. (2011). A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export, Expert Systems with Applications 38:9296–9304.

14. Zhang T., Zhang X., Liu Y., Luo Y., Zhou T., and Li X. (2016). The analysis of infectious disease surveillance data based on fuzzy time series method, International Journal of Infectious Diseases, 45, 309-310.

15. Tricahya S. and Rustam Z. (2019, June). Forecasting the Amount of Pneumonia Patients in Jakarta with Weighted High Order Fuzzy Time Series. In IOP Conference Series: Materials Science and Engineering (Vol. 546, No. 5, p. 052080). IOP Publishing.

16. Akaike H. (1974) A New Look at the Statistical Model Identification, IEEE Transactions on Automatic Control, AC- 19, 716-723.

17. Song Q., Chissom BS (1994) Forecasting enrollments with fuzzy time series—part II, FuzzySets Syst 62(1):1–8.

18. Zadeh L.A. (1965). Fuzzy sets, Inf Control 8(3):338–353.

19. Chen S.M. (1996). Forecasting enrollments based on fuzzy time series, Fuzzy Sets Syst 81:311–319.

20. Huarng K. (2001). Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets Syst 123:387–394.

Figures

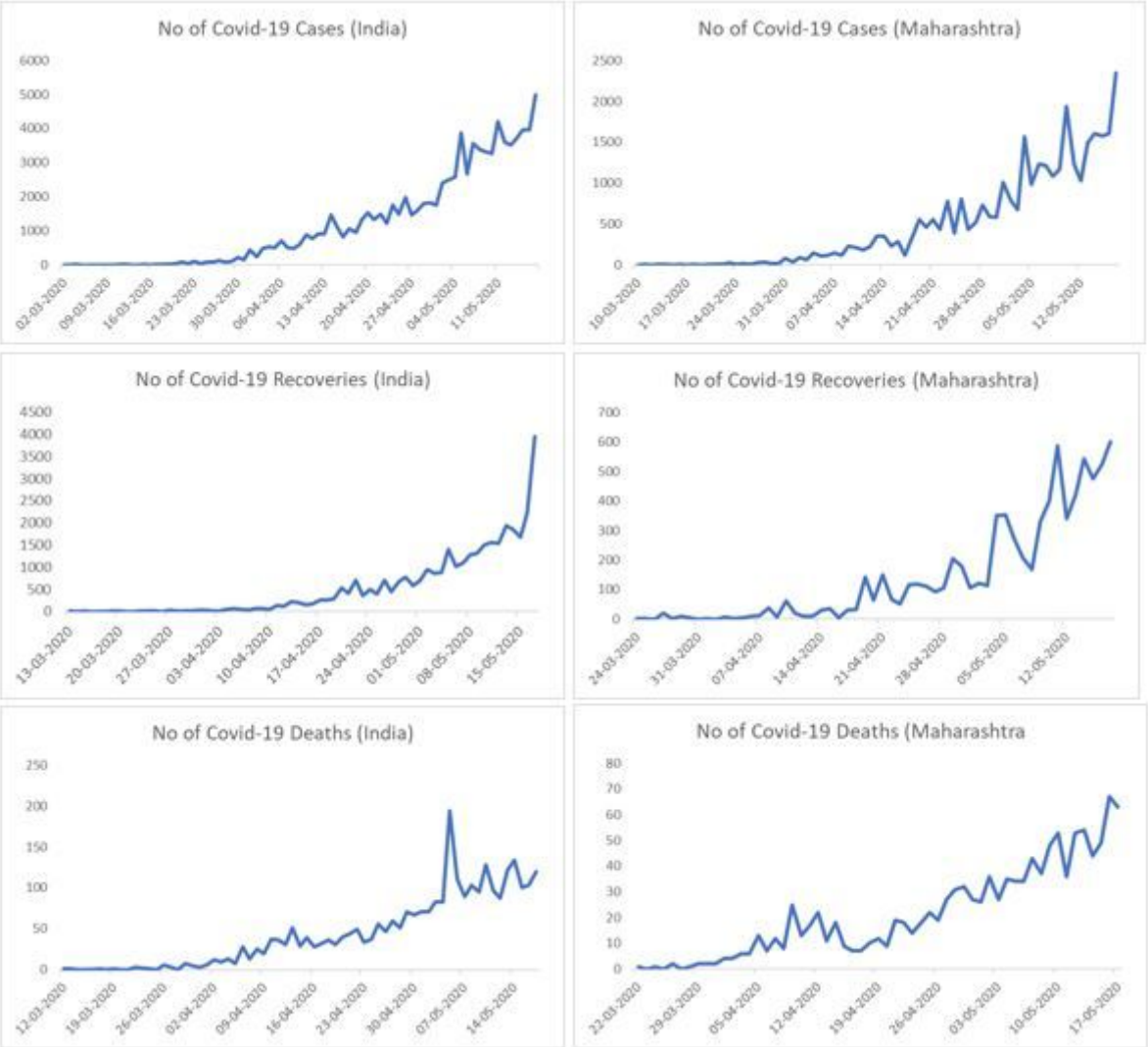


Figure 1

Graphical Representation of COVID-19 statistics in India and Maharashtra

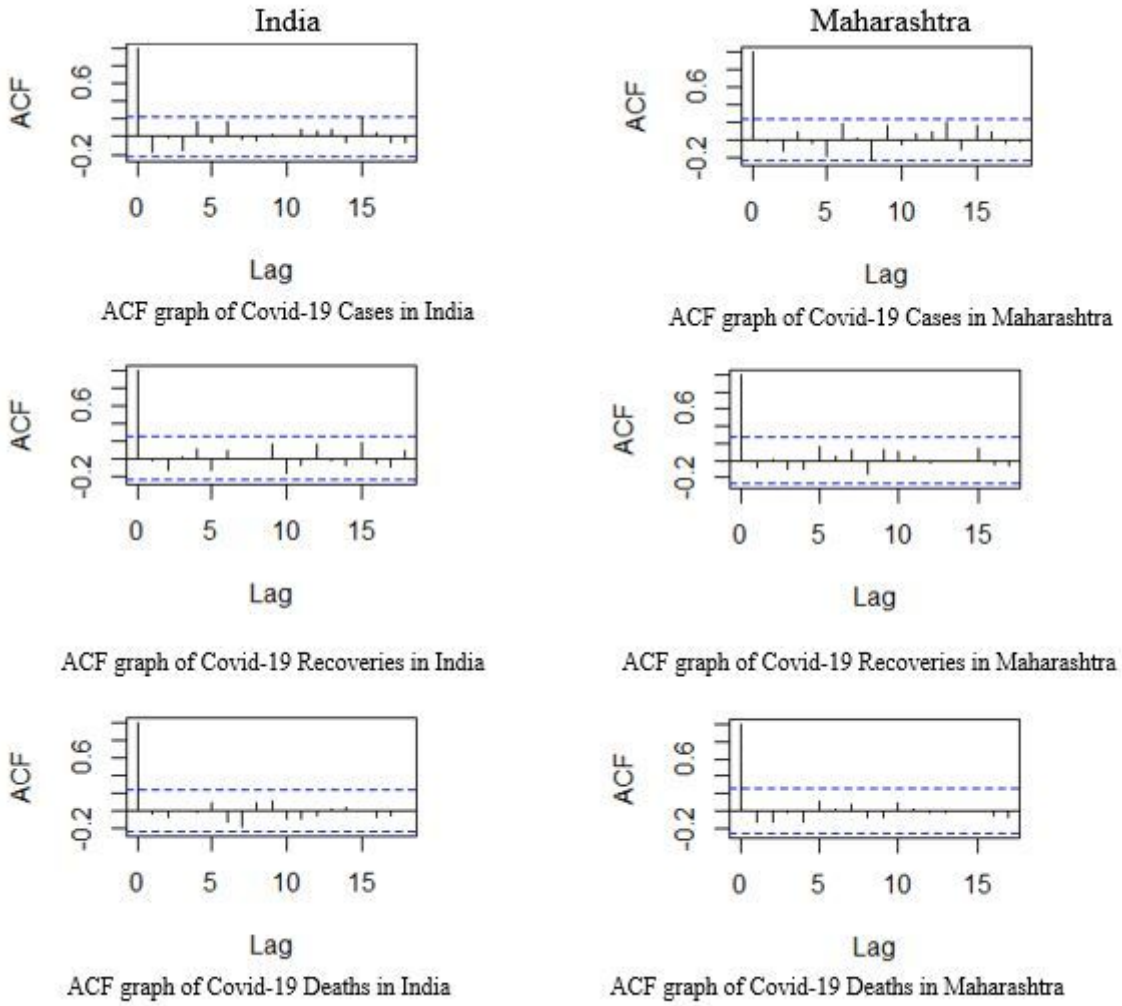


Figure 2

Graphical representation of residual ACF of COVID-19 data samples

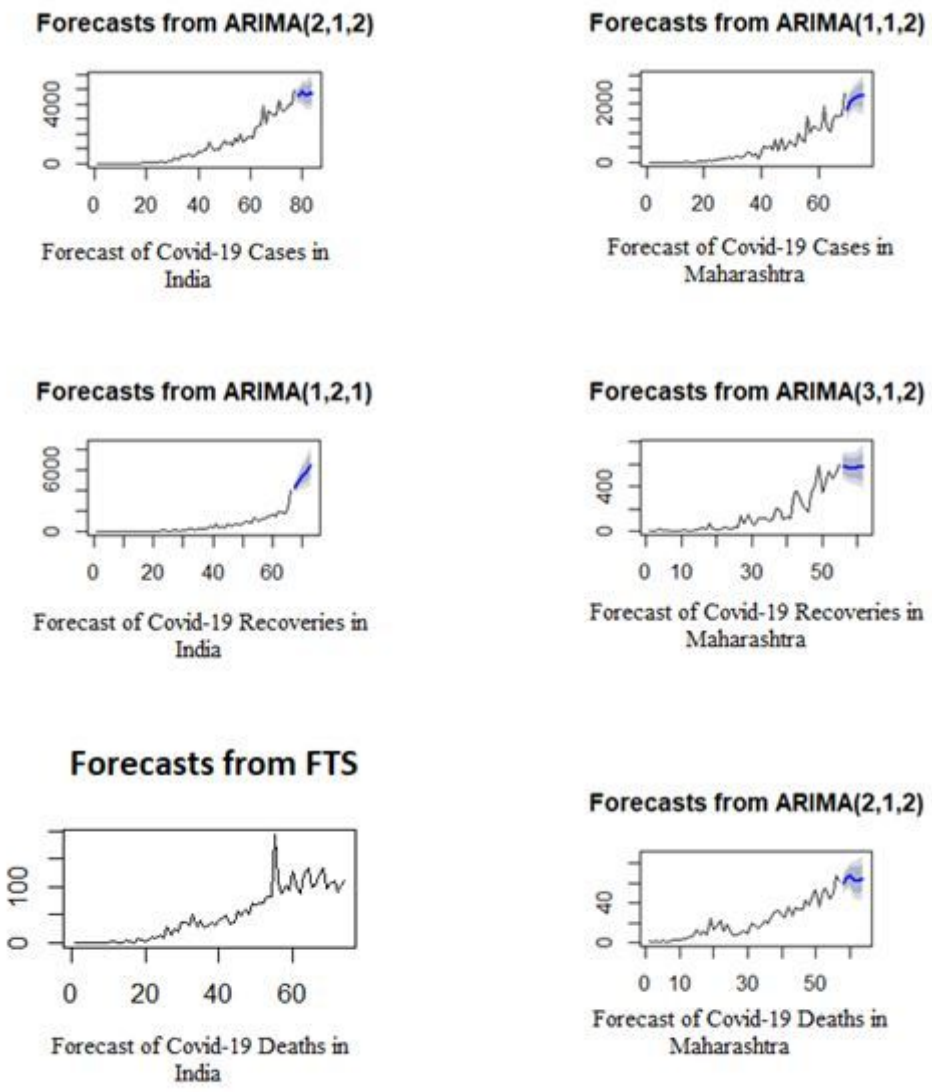


Figure 3

Graphical representation of Forecasted values of the COVID-19 pandemic with ARIMA and FTS Model

Figure 4

Comparison between the actual and forecasted values of COVID-19 data sets