



A Spreadsheet-Based Short Time Forecasting Method for the COVID-19 Pandemic

Ritam Pal¹ · Sourav Sarkar¹ · Achintya Mukhopadhyay¹

Received: 19 May 2021 / Accepted: 13 August 2021 / Published online: 17 August 2021
© Indian National Academy of Engineering 2021

Abstract

As we are writing this paper, the number of daily affected COVID patients is around 0.38 million and with active cases over 3 million in India. This large number of active cases is putting the medical facilities under severe strain. Many researchers have proposed many ways of forecasting the COVID-19 patients but they mainly worked on the cumulative cases and moreover, all those methods required considerable skill and computational cost. In this work, a simple spreadsheet-based forecasting model has been developed which will help to predict the number of active cases in the immediate future i.e., the next few days. This information can be useful for emergency management. The difficulty which is generally faced in predicting the active cases is that the dynamics of active cases has a complex dependence on a number of Non-Pharmaceutical Interventions (NPI) and social factors and can undergo sharp changes. Quadratic, cubic and quartic polynomial functions have been applied to capture these peaks and observed that the quadratic function helps in better prediction of the peak. The accuracy of the prediction methods is measured as well as it is tried to observe how the methods predict data for the next 1 day, 3 days and 6 days. A prediction method analogous to weather forecasting method is recommended in this work where the prediction for each day gets updated depending on the most recent data available. This method has also been found to perform well even in the period there were sharp changes in the trend due to imposition of strict NPI measures.

Keywords Spreadsheet · Quadratic · Cubic · Active cases · Prediction

Introduction

The COVID-19 pandemic has already created havoc in the entire world and it is still affecting the health of millions of people worldwide including India. In addition to health issues, the pandemic has caused a lot of distress in the global economy. Implementation of Non-Pharmaceutical Intervention (NPI) is important to control the spread of the disease. NPI includes the utilisation of face masks, hand sanitizers, application of quarantine rules and the imposition of partial lockdowns by the respective governments (Bell et al. 2006). The implementation has to be carried out at the right time otherwise it will become very difficult to reduce the spread of the SARS-CoV-2. To carry out proper NPI implementations, some prediction and forecasting will help the policy makers in assessing the impact of different NPI measures on

the pandemics progression. However, NPI measures yield benefits over relatively extended time spans and their benefits may be observed only after some time. Hence these models are more suited for framing policies like mass vaccination, travel restrictions, partial or complete lockdowns over intermediate time ranges spanning several weeks to a few months. Several long-term forecasting models have been prescribed and reported in different literatures. Giordano et al. proposed the SIDARTHE model to define the reproduction number in Italy (Giordano et al. 2020). Wangping employed the Markov Chain Monte Carlo method to predict the reproductive number in Italy and Hunan (Wangping et al. 2020). Some studies involved the use of mathematical models to estimate the loss of human lives and death count. The models employed are susceptible-infected-recovered (SIR) model and Boltzmann's function-based analysis (Botha and Dednam 2020; Gao et al. 2020; Lana et al. 2020; Magal and Webb 2020; Okhueuse 2020; Sameni 2020; Wang et al. 2020; Weber et al. 2020). There are many works where long short-term memory (LSTM) model have been used to forecast the pandemic data (Chimmula and Zhang 2020; Komarova et al.

✉ Achintya Mukhopadhyay
achintya.mukhopadhyay@jadavpuruniversity.in

¹ Department of Mechanical Engineering, Jadavpur University, Kolkata 700032, India

2020; Tomar and Gupta 2020; Alvarez et al. 2021; Guleryuz 2021). Some trend and analysis have been studied in India based on SIR model (Sarkar et al. 2020). Several works have reported methods of prediction of cases which are influenced by lockdown or interventions. Verma et al. proposed a power law curve between the growth phase and the flattening phase, which takes interventions into account (Verma et al. 2020). Bhardwaj employed a regression analysis with least square fitting to establish a logistic model that predicted the cases during restricted movement (Bhardwaj 2020). Ranjan utilized Gaussian distribution with piecewise linear fitting to predict the number of cases influenced by lockdown (Ranjan 2020). All these works focussed on the total number of cases but the active number of cases inform the authorities regarding the requirements of hospital beds and the necessary items. In addition, these works focussed on dynamic trends over extended periods of time.

On the other hand, several disaster management measures, including public health related issues, may have to be implemented at very short notice. For successful implementation of these measures, accurate predictions, valid over very short time intervals, are necessary. This is in contrast with the long-term projections normally needed for formulation of different NPI policies over medium to extended periods of time. As the spread of the disease will grow, the number of patients who are under the treatments at various health care facilities will grow and proper planning would be needed for emergency arrangement of adequate human and material resources for crisis management. A prior estimate of the expected number of cases would be of great help to the authorities. Short-term forecasting models help the policy makers to understand the amount of challenge which they have to face in the immediate future, say, the coming week. Many researchers have done prediction based on the available data sets and employed the best Machine Learning (ML) model corresponding to the dataset (Batista 2020; Fong et al. 2020; Hu et al. 2020; Jia et al. 2020). A short-term forecasting of cumulative COVID-19 cases in Brazil with one, three and six days ahead have been done based on Auto-regressive Integrated Moving Average (ARIMA), cubist regression (CUBIST), random forest (RF), ridge regression (RIDGE), support vector regression (SVR) and stacking ensemble learning (Ribeiro et al. 2020). A work has been reported using US county dataset where the forecasting has been done based on a spatio-temporal Neural Network Approach (Kapoor et al. 2020). In India, some short-term forecasting models have been reported based on ARIMA (Gupta and Pal 2020; Ganiny and Nisar 2021). Some major disadvantages of the ARIMA method are that the method is subjective and the reliability of the model depends upon the skill and experience of the forecaster (Yusheng 2005). Moreover, the method is computationally expensive and the underlying theoretical model and structural relationships are

not distinct. The model adds some complexity to the forecasting. Big data analysis techniques and ML techniques are expensive in terms of computation and skilled personnel are required to implement these techniques.

Although a large volume of literature has been produced that deals with trend analysis and prediction of the dynamics of COVID-19 pandemic in different parts of the world, some of the issues, listed below, have received relatively less attention compared to the effects of different epidemiological and NPI measures on long term progression of the pandemic. The novelty of the work, compared to the majority of works reported can be summarised as follows.

- Compared to most other works, this work is concerned about prediction over very short intervals of time using data from the immediate past. While this technique may not be sufficient to guide long or medium time policy making, these predictions can be of immense use to emergency planning, especially related to healthcare services. As the results show, the characteristics of the short time series considered here vary considerably from the moderate to long time series considered in most other works.
- Most of the works based on time series analysis related to COVID-19 focussed on the cumulative cases or daily infections. Very few of these works predicted active cases. Prediction of active cases has its own challenges as it is determined by the competing effects of new infections and recoveries, each of which is governed by different sets of parameters.

The major motivation of this work is to develop a simple, computationally inexpensive model that can be used for accurate short-time forecast through regular updating of available data. This procedure can be applied at local, regional and national levels and requires only the most basic data, which makes it suitable for easy implementation at all levels for disaster management.

In this work, a simple spreadsheet-based forecasting method is proposed for the COVID-19 cumulative cases as well as the active patients who are undergoing treatment in the hospitals. The forecasting method has been applied on data of patients in India as a whole and also on two states, West Bengal and National Capital Region of Delhi. The method involves fitting a polynomial function to the available data set and based on the polynomial, prediction for next 1 day, 3 days and 6 days have been carried out. Quadratic, cubic and quartic polynomials have been tried out. It is tried to predict both cumulative as well as active cases. For planning healthcare facilities, it is the active cases that deserve more attention. Before going on to prediction, the polynomial functions are fitted on to the actual dataset and the correlation coefficient values are compared. While

predicting the active cases, some peaks and inflection points are encountered which are challenging issues for prediction. The polynomial functions helped in the forecasting and the errors for each polynomial function are compared to conclude which polynomial function gives best result. The change in slope or inflection becomes important for the active number of patients. So, the aim is to forecast the active cases more accurately as this will help the policy makers to better understand the scenario and get prepared. The method has also been used for cumulative or total number of cases and the results show an excellent match with the actual data. We would like to emphasize that the present work is not a conventional epidemiological model and hence direct assessment of the impacts of different epidemiological and NPI measures on the pandemic dynamics is beyond the scope of the present work. The focus of the work is analysis of trend of the time series to predict the cumulative and active cases over short time intervals. However, the model is found to be robust enough to successfully capture sudden changes in trend due to implementation of strict NPI measures including partial lockdown and the predictions are well in coherence with the actual numbers.

Methods

A theoretical function is assumed which represents the epidemic count with time in a country or region, as follows,

$$y = f(x)$$

where y represents the number of cumulative cases or the number of active cases and x represents the corresponding day for which the numbers are calculated. The function has been assumed to be a polynomial. It is based on

the observed capacity of the polynomial to fit accurately in longer or shorter sections of an epidemic curve. Historically, it has been seen in many cases that an epidemic relation follows a normal distribution and a cubic polynomial fits very well in that distribution barring the start and end (Susser and Adelstein 1975; Orihuel et al. 2021). But in these former works, they have used datasets for cumulative cases. Here, both cumulative and active cases dataset are being used. Therefore, both active and cumulative cases are modelled as quadratic, cubic, and quartic cubic polynomial functions of time as follows,

$$y = ax^2 + bx + c$$

$$y = ax^3 + bx^2 + cx + d$$

$$y = ax^4 + bx^3 + cx^2 + dx + e$$

Without NPI implementations, the cubic polynomial proves to be an accurate method in modelling and forecasting the pandemic (Orihuel et al. 2021). A curve fitting is presented in Fig. 1 showing how the polynomial curves fits with the actual data from West Bengal, a state in India with Fig. 1a showing the cumulative number of cases and Fig. 1b showing the active number of cases. The dataset which has been utilized in this work is obtained from this website (<https://prsindia.org/covid-19/cases>). From the curves, it can be seen that in both the cases, the quadratic and the cubic polynomials closely follow the actual dataset which is evident from their correlation coefficient values. In both the situations, the polynomials fit with R^2 above 0.99. R^2 denotes the square of the correlation coefficient which is found by regression analysis and the curve fitting has been done using the least-square residual method (Schlossmacher 1973; Rodgers and Nicewander 1988). Results with quartic

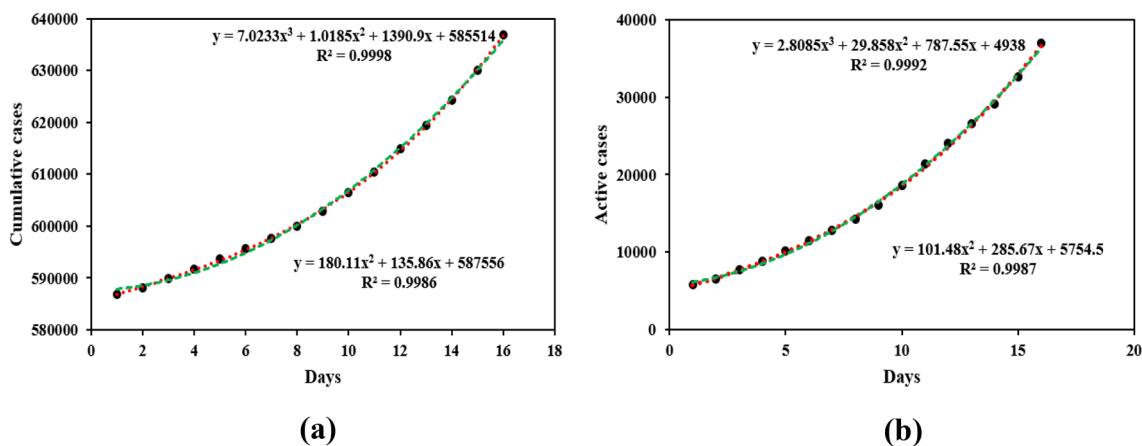


Fig. 1 Quadratic (green dashed) and Cubic (red dashed) polynomial fitted into actual data of **a** cumulative cases and **b** active number of cases from West Bengal, India; black solid-actual data points; equation and R^2 values shown within the pictures

polynomial have been presented in the discussions of error values. The authors have mainly focussed on cubic and quadratic functions as we will see in the later sections that quartic polynomials result in an appreciable amount of error.

With this available data set, predictions and forecasting have been carried out. The prediction data are built up as follows. Let us suppose, there is a dataset from 1st to 7th day. A polynomial curve (quadratic, cubic or quartic) is used to fit the data using least square curve fitting technique. The fitted polynomial function is then used to predict the values for the next 6 days i.e., 8th to 13th day. Again, using the dataset of 2nd to 8th day (available on and after 8th day), prediction is carried out for 9th to 14th day. In this way, for each day predictions are continuously updated based on the most recent trend. This case is analogous to meteorological forecasting for natural calamities like cyclone where the forecasting keeps on getting updated at regular intervals depending on the most recent weather data available. This technique allows the method to adjust the predictions to even very sharp changes and reversals in trend very quickly. As there are predicted values for $n + 6$ days, it has been tried to calculate how accurately the prediction works without being updated. This means that for each data window of 7 days, we are having prediction for $n + 6$ days, where n denotes the last day of the data window and we are measuring the prediction accuracy for $(n + 1)$ th day, $(n + 3)$ th day and $(n + 6)$ th day. A flowchart giving an outlook on the prediction method is displayed in Fig. 2. To obtain the accuracy, the deviation of the predictions for those days from the actual values of the corresponding days is being measured. The accuracy of the prediction is quantified in terms of the following parameters:

$$APE = \frac{|y_{actual} - y_{pred}|}{y_{actual}}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_{i(actual)} - y_{i(pred)}|}{y_{i(actual)}}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i(actual)} - y_{i(pred)})^2}$$

where, y_{actual} represents the actual value for a given day and y_{pred} denotes the predicted value for that particular day, i represents the particular predicted day, n represents the last day of prediction range, APE is Absolute Percentage Error, MAPE denotes Mean Absolute Percentage Error and RMSE implies Root Mean Squared Error.

Results and Discussions

Comparison of Predicted and Actual Cases

The prediction method (based on data set updated with the latest information) is employed to build the predicted dataset to observe how well the prediction follows the actual scenario. Firstly, the cumulative cases of Delhi are used to verify whether the method is working properly. In Fig. 3a, b, it is investigated how well the predicted values follow the actual curve and the error associated with those predictions respectively. The quadratic polynomial function results in a higher error value than the cubic polynomial function throughout the data range. The error values tend to rise when there is presence of inflection or sharp rise. From this section, it can be concluded that cumulative cases can be fitted by cubic functions more accurately and also, it has been historically observed (Susser and Adelstein 1975).

It has been seen that the prediction of cumulative cases is very accurate in terms of error and the prediction method is extended to active number of cases. To predict active cases, two datasets from West Bengal and New Delhi are used. In the West Bengal dataset (Fig. 4a), there is presence of only one peak due to which the predicted values have followed the actual data with great accuracy. When one looks into the data from Delhi (Fig. 4b), there are a number of inflection points and three peaks which require the predicted data to undergo repetitive corrections and follow the actual data curve, which is observed from the cluster of predicted values near the peak regions. In case of Delhi, it can be seen that the cubic function deviates from the actual curve near the peak much more than the quadratic function.

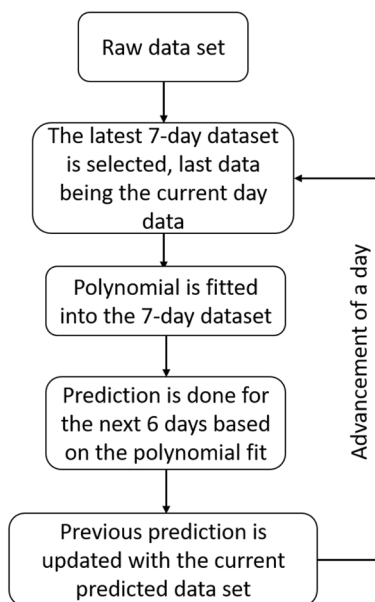


Fig. 2 Flowchart showing the way of prediction

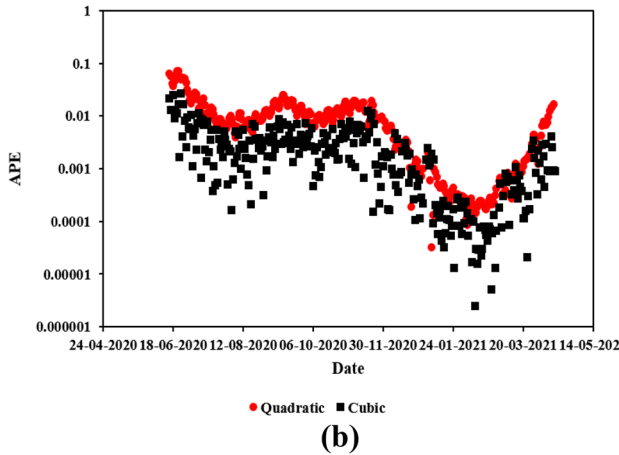
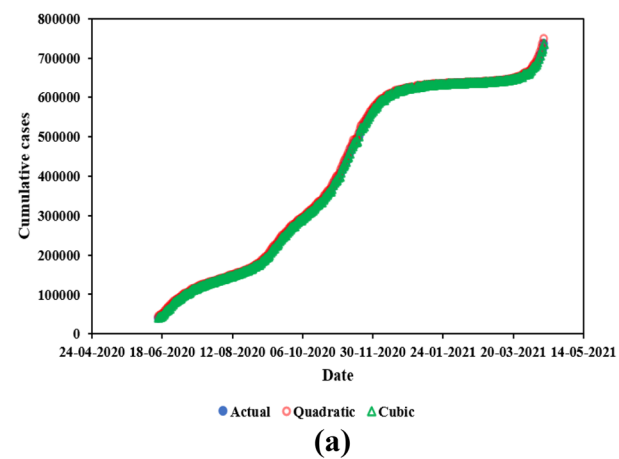


Fig. 3 **a** Prediction of cumulative cases in New Delhi and **b** error associated with the prediction methods

Procedure for Predicting Sharp Changes in Trend and Inflection

To capture the inflection points more accurately, the dataset range is gradually shifted in the forward direction and the cases for the next six days are predicted with both quadratic and cubic polynomial functions. The exercise is described as follows and the dataset of West Bengal is utilized to carry out the prediction of the peak. At first, the actual dataset was up to 25th October 2021 and we predicted the cases till 31st October (Case 1). The quadratic function has sensed the presence of peak (Fig. 5a) and the predicted values show a plateau type zone whereas the cubic function is unable to sense that peak and the predicted values went on increasing (Fig. 5b) and overestimating the actual peak. Then, the dataset was up to 27th October and the cases were predicted till 2nd November (Case 2). In this case, the predicted values from the quadratic function show an agreement with the actual data but the cubic polynomial underestimates the prediction. The next dataset was till 31st October and the prediction was done up to 6th November (Case 3). There

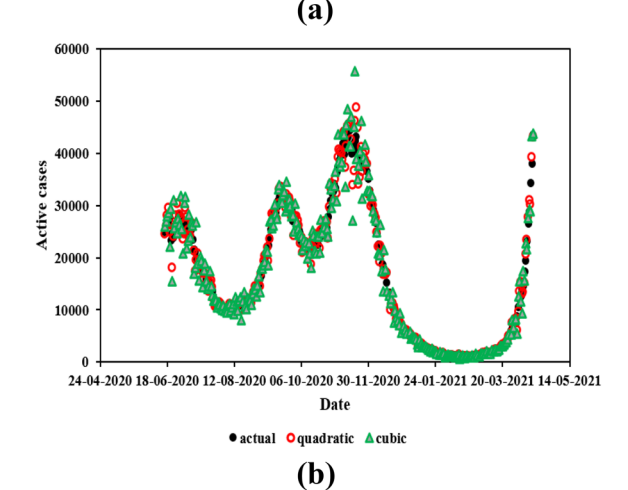
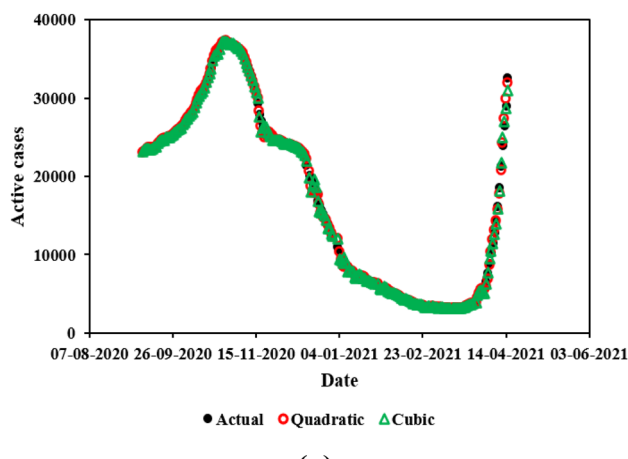


Fig. 4 Actual and predicted dataset for **a** West Bengal and **b** New Delhi in India

is a small amount of fall in the quadratic predicted values while the cubic predicted values follow a rise steeper than the actual dataset range. Finally, the predicted values for both the polynomials became more or less coherent with the actual values when the dataset was up to 2nd November and the prediction was till 8th November (Case 4). From this exercise, one finds that the quadratic polynomial is better in sensing and forecasting a peak than the cubic polynomial. All the cases are presented in Fig. 5 where prediction with quadratic is presented in Fig. 5a and cubic is presented in Fig. 5b.

In a similar manner, the behaviour of the polynomial functions is shown when a sharp rise in active cases is encountered. The predictions from quadratic polynomial are presented in Fig. 6a and those from cubic polynomial are shown in Fig. 6b. The dataset for prediction has been shifted to predict the sharp rise more accurately. Case1 corresponds to dataset from 1st to 7th April, Case2 corresponds to dataset up to 9th April, Case3 contains dataset up to 12th April and Case4 is up to 15th April. In all these cases, the

prediction has been done for the next 7 days. For Case1, the cubic polynomial fails to capture the rise, instead the prediction attains a peak and falls but the quadratic polynomial predicts the rise. For Case2 and Case3, both the polynomials try to predict the cases with some error. Case4 shows that the cubic polynomial overestimates the rise but the quadratic polynomial predicts values close to the actual data. A dataset has been provided in Table 1 to show the changes in prediction which is caused the shifting of the dataset. From the

table one can see how the prediction corrects itself to follow the actual dataset. Thus, quadratic polynomial proves to be better in case of sharp rises in active cases.

Calculation of Error

So far, the methods of prediction have been established. Now time has come to focus on the deviation of the predicted values from the actual values. The prediction is

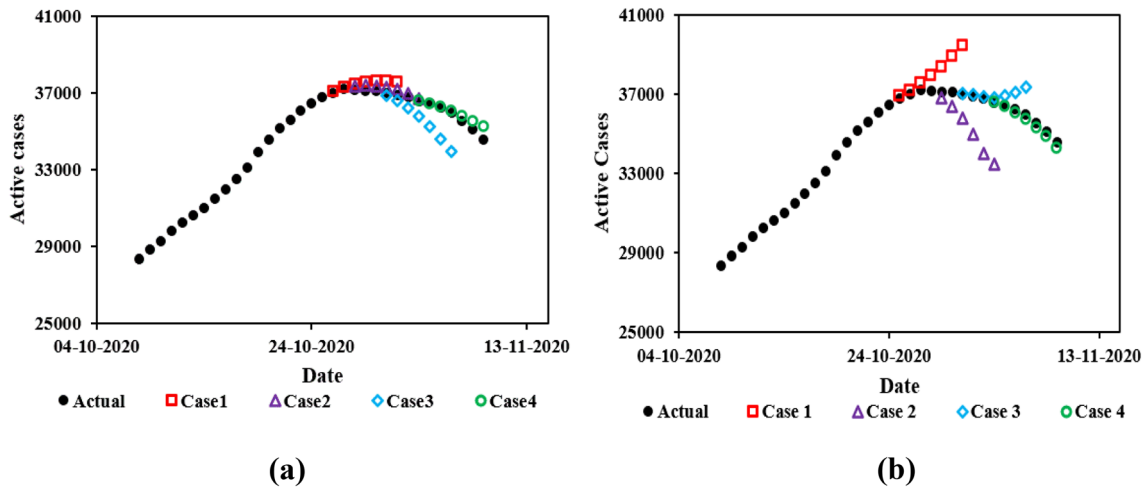


Fig. 5 Shifting of dataset to capture or predict the peak more accurately; a quadratic, b cubic

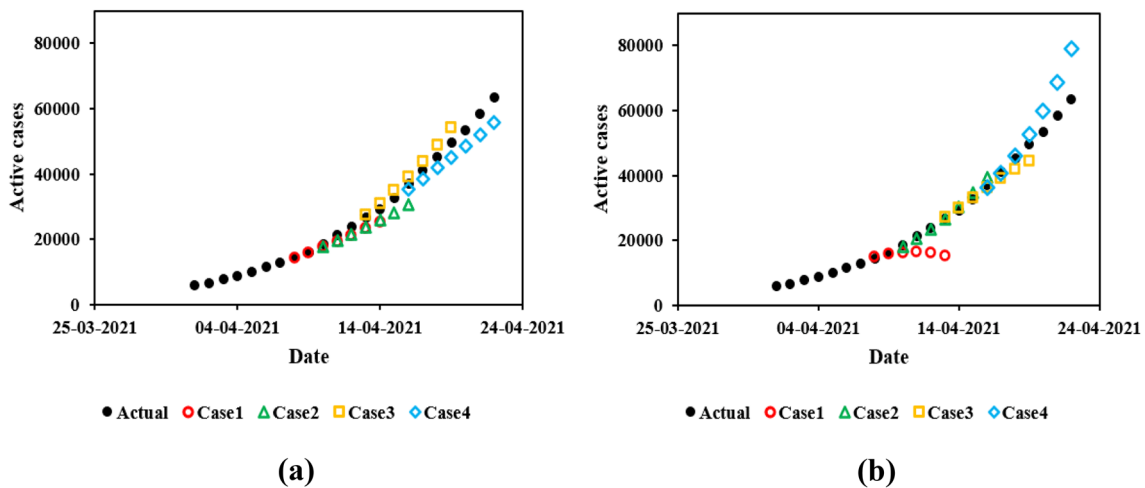


Fig. 6 Way of predicting a sharp rise in active cases; a quadratic function, b cubic function

Table 1 Way of changing the prediction by shifting the dataset in the forward direction by using quadratic function

Date	Actual active cases	Prediction by shifting dataset				Mean APE
		Case1	Case2	Case3	Case4	
14-04-2021	29,050	25,318	25,860	30,997	–	0.10
16-04-2021	36,981	–	30,531	39,224	35,413	0.09

done for the next 6 days and it is inevitable that the predictions will vary from the actual data. The range of variation is important and to understand the range of variation, we are calculating the error in the predicted values for $(n + 1)$ thday, $(n + 3)$ thday and $(n + 6)$ thday. From these error values one will be able to understand whether the method of updating the prediction is necessary or not. As the active cases in New Delhi consist of a number of peaks, we are using that dataset from June 2020 to April 2021 to obtain the range of error in the prediction. The errors for the 3 respective days are shown in Fig. 7. Error obtained from quadratic, cubic and quartic polynomial functions are shown in the figure. The error in prediction shown for $(n + 1)$ th day in Fig. 7a reveals that the maximum error is around 0.45 (or 45%) for both cubic and quartic but for quadratic, it is around 0.1 (10%). The minimum error is around 0.0005 (or 0.05%) for all the polynomials. The maximum error is occurring around the dates where there is presence of inflection zones. Following the error for $(n + 1)$ th day, error values for $(n + 3)$ th day and $(n + 6)$ th day are presented in Fig. 7b, c respectively. In Fig. 7b one can notice an increase in the error values compared to Fig. 7a and the error values further increased in Fig. 7c but the quadratic polynomial is always showing a less error compared to the other polynomial functions. This clarifies that there is necessity in using the latest predicted values for a day's prediction and it will be close to the actual value and for a range of 6 days prediction, quadratic polynomial proves to be a better tool than cubic polynomial.

There is another way of representing the error associated with the prediction methods as well as with the range of predicted values. When values for the next 3 days are predicted, there will be 3 predicted values for each day. So, for a particular day, there will be maximum, minimum and mean error. If we are predicting for next 6 days, there will be 6 values for a particular day and similarly, there will be maximum, minimum and mean error associated with that day. From this presentation, one will be able to observe how the associated errors change with the range of predicted values and how these methods are effective for short term forecasting. In Fig. 8a, b, we have shown the maximum, minimum and mean APE values for the prediction range of $n + 3$ days and $n + 6$ days with quadratic and cubic functions. It is observed that the entire range of error values get shifted upwards or in other words the error increases when one changes the prediction range from $n + 3$ to $n + 6$ days. For $(n + 3)$ days, both the polynomial functions occupy the same zone for the APE values with some APE values for quadratic function is showing less error at some points. When prediction is done for $(n + 6)$ days, there is a significant difference in the APE values for quadratic and cubic functions. The quadratic function results in less error for a long range of prediction compared to cubic polynomial.

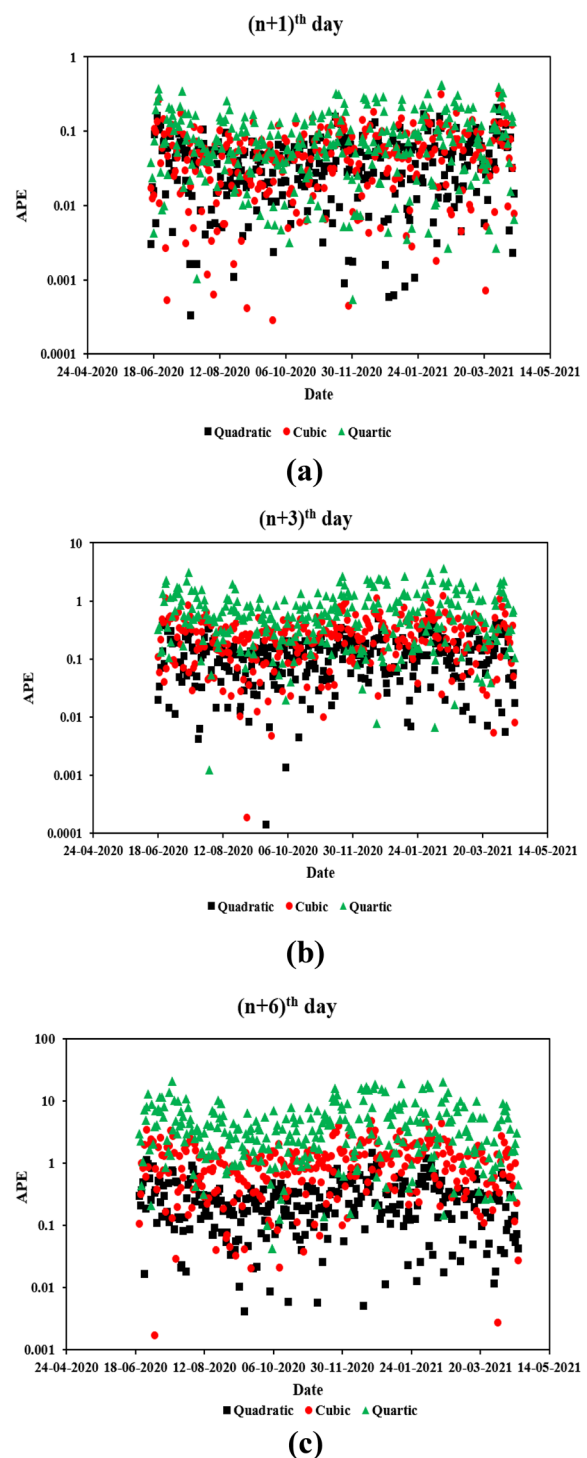


Fig. 7 Error in predictions for **a** $(n + 1)$ th day, **b** $(n + 3)$ th day and **c** $(n + 6)$ th day

Some more types of errors are tabulated for our different prediction approaches and for different prediction days in Table 2. All the errors have been calculated for active cases of New Delhi for the time period of June 2020 to April 2021. From the Table 2, one can understand that

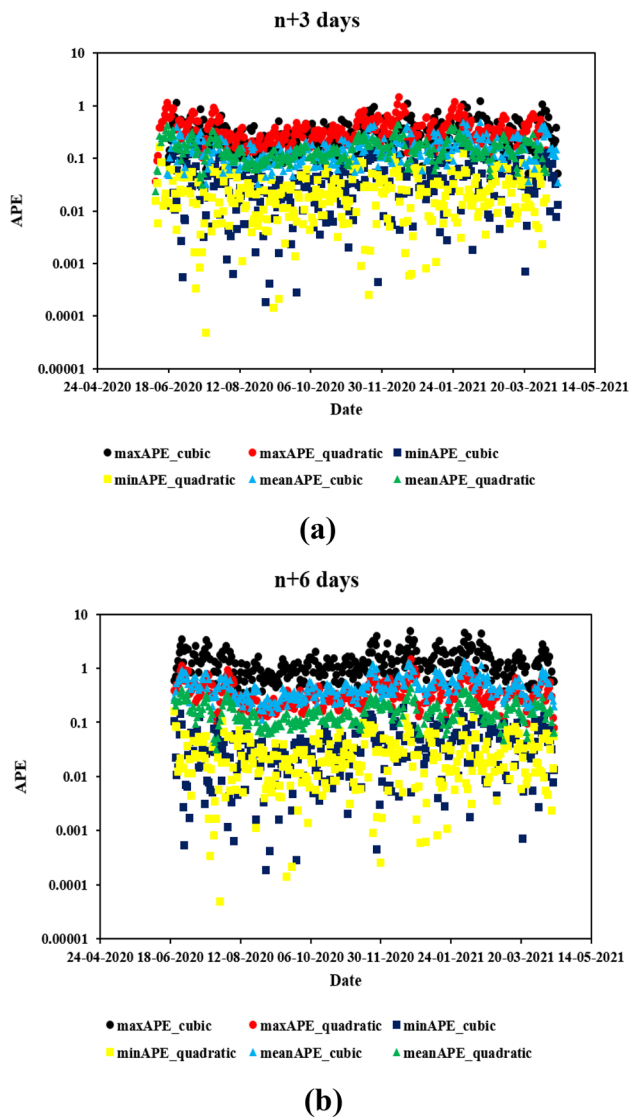


Fig. 8 Maximum, minimum and mean error for each day’s predicted value with different prediction ranges—**a** $n + 3$ days and **b** $n + 6$ days

the active cases can be better predicted by the quadratic polynomial. Even for the prediction of six future days, MAPE is 0.28 (or 28%). If we look into cumulative cases prediction for $(n + 1)$ days, MAPE for cubic polynomial is 0.003 whereas for quadratic polynomial MAPE is 0.0099.

Table 2 Comparison of different types of errors for different prediction methods and ranges

		Day for which prediction is made					
		(n + 1)th day		(n + 3)th day		(n + 6)th day	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Prediction method	Quadratic	0.044	1147	0.13	3300	0.28	7384
	Cubic	0.059	1540	0.277	7300	1.1	27,873
	Quartic	0.094	2494	0.73	19,424	4.5	116,280

Thus, for cumulative cases, cubic polynomial gives better prediction.

A prediction of cumulative and active cases is provided along with the APE values. Prediction is done for 7 days for cumulative as well as active cases employing cubic polynomial for cumulative and quadratic for active cases. Data up to 25th April has been used for the prediction and one can see a very good agreement with the actual data in the two categories. Prediction for West Bengal as well as India is presented in Table 3a, b respectively. The predictions are also shown in Fig. 9. A coherent prediction can be concluded by taking a look at the APE values. The maximum APE value is 19.6% for the 7th day prediction of active cases and 1.4% for the total cases in West Bengal. The corresponding figures are 11.3% and 2.8% for all India data.

Impact of NPI

As stated earlier, direct assessment of different epidemiological and NPI measures on the progression of infection is not the focus of the paper. However, implementation of strict NPI measures (e.g., lockdown like restrictions) and major changes in epidemiological factors like virulence of new strains of virus bring about sudden changes in number of infected and active cases that have immediate impacts on public health measures. So, it is important to assess the capability of the model to such bifurcation points in the pandemic dynamics. In order to investigate the robustness of the present model to sudden changes brought about by the impact of interventions, cases during the lockdown periods have been considered. In West Bengal, the government imposed comprehensive restriction on public movement from 16th May 2021. Limited relaxation was announced from 1st June 2021 followed by partial relaxation from 16th June 2021. In this period, the cumulative as well as the active cases have been influenced by the restrictions and with the proposed model, the cases during these periods are predicted. In Fig. 10, the way of predicting the steep decrease in active cases has been shown from the start of the restriction period. The figure also includes the period of limited as well as partial relaxation and how the quadratic polynomial tries to predict the slight rise in cases after the steep fall in numbers. The

Table 3 a Prediction for 7 days in West Bengal along with the actual available data. **b** Prediction for 7 days in India along with the actual available data

a						
Date	Cumulative cases			Active cases		
	Actual	Predicted	APE	Actual	Predicted	APE
26-04-2021	743,950	743,355	0.000799	88,800	88,234	0.006369
27-04-2021	759,942	759,698	0.00032	94,949	95,621	0.007086
28-04-2021	776,345	777,137	0.00102	100,615	103,474	0.028418
29-04-2021	793,552	795,666	0.002665	105,812	111,791	0.056512
30-04-2021	810,955	815,284	0.005338	110,241	120,574	0.093731
01-05-2021	828,366	835,984	0.009197	113,624	129,821	0.142552
02-05-2021	845,878	857,763	0.014051	116,659	139,533	0.196081

b						
Date	Cumulative cases			Active cases		
	Actual	Predicted	APE	Actual	Predicted	APE
26-04-2021	17,313,163	17,308,822	0.0002508	2,813,658	2,826,296	0.004492
27-04-2021	17,636,307	17,649,425	0.0007438	2,882,204	2,968,509	0.029944
28-04-2021	17,997,267	17,975,564	0.0012059	2,978,709	3,113,965	0.045408
29-04-2021	18,376,524	18,280,787	0.0052097	3,084,814	3,262,664	0.057653
30-04-2021	18,762,976	18,558,645	0.0108901	3,170,228	3,414,607	0.077086
01-05-2021	19,164,969	18,802,685	0.0189034	3,268,710	3,569,793	0.092111
02-05-2021	19,557,457	19,006,458	0.0281733	3,349,644	3,728,223	0.113021

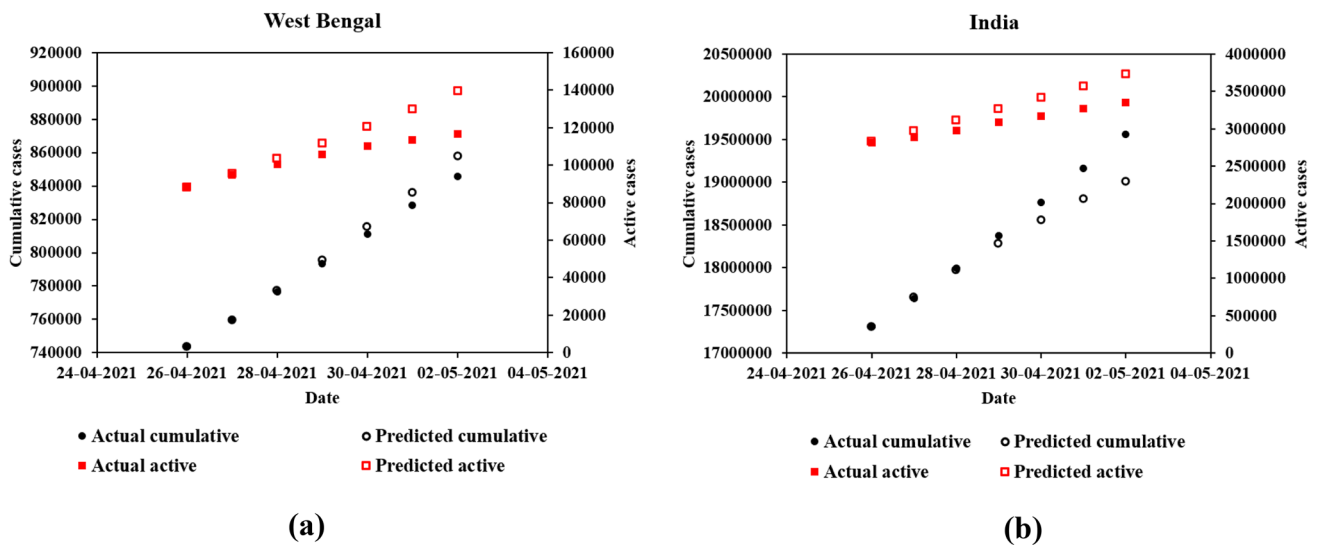
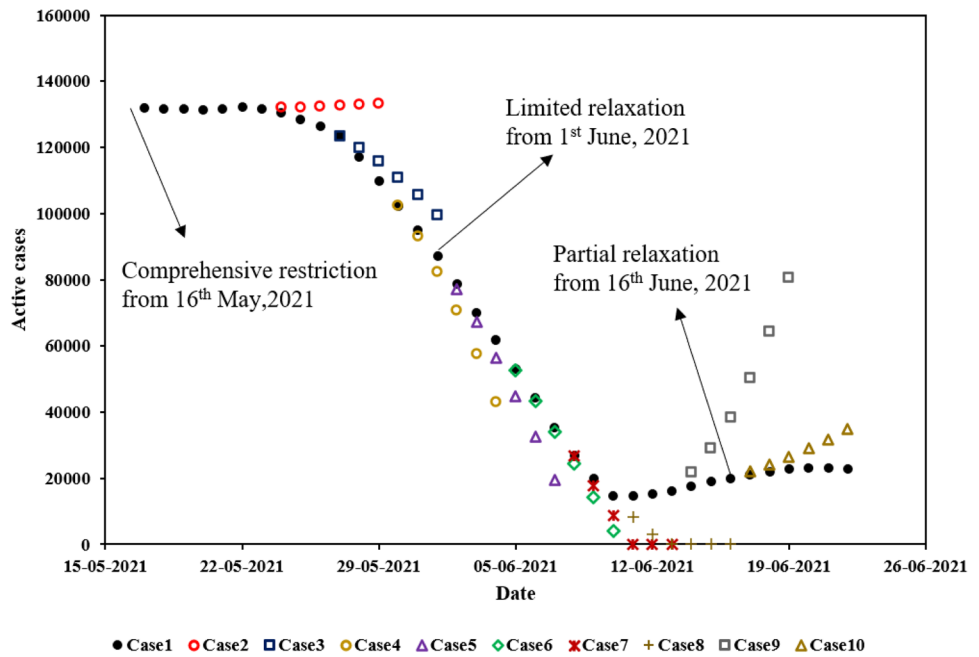


Fig. 9 Predictions with actual data for cumulative as well as active cases for 7 days in **a** West Bengal and **b** India

prediction has been done in the similar fashion as mentioned for an inflection point like steep rise or a peak. A 7-day data range is selected and prediction is done for next 6 days and the data range is shifted in the forward time direction. Figure 10 consists of 10 cases of date ranges: Case1- 17th May to 23rd May 2021, Case2- 20th May

to 26th May 2021, Case3- 23rd May to 29th May 2021, Case4- 26th May to 1st June 2021, Case5-29th May to 4th June 2021, Case6- 1st June to 7th June 2021, Case7- 4th June to 10th June 2021, Case8- 7th June to 13th June 2021, Case9- 10th June to 16th June 2021, Case10- 13th June to 19th June 2021. The predictions show that the method at

Fig. 10 Showcasing the predictions in the periods of interventions; Case1- 17th May to 23rd May 2021, Case2- 20th May to 26th May 2021, Case3- 23rd May to 29th May 2021, Case4- 26th May to 1st June 2021, Case5-29th May to 4th June 2021, Case6- 1st June to 7th June 2021, Case7- 4th June to 10th June 2021, Case8- 7th June to 13th June 2021, Case9- 10th June to 16th June 2021, Case10- 13th June to 19th June 2021



first predicted a constant number of active cases after the first restriction imposition but the method quickly adjusted to the steep fall in the active number of cases. When the limited relaxation happened, the method at first predicted a further decrease in the number of cases based primarily on the immediate history during the period of comprehensive restriction but a rise in the actual cases occurred which forced the method to initially overpredict the number of cases and then the method tried to predict close to the actual number of cases. So, the steep fall followed by a

slight rise in the number of cases caused the method to under predict and over predict but the method followed the trend quickly. In Fig. 11, both cumulative and active cases during these periods have been predicted using the cubic and quadratic polynomial methods with regular data updation (similar to tracking and predicting meteorological catastrophe like cyclones). In both the cases, the prediction is coherent with the actual cases which implies that the influence of interventions can also be well captured by the proposed methods in this work.

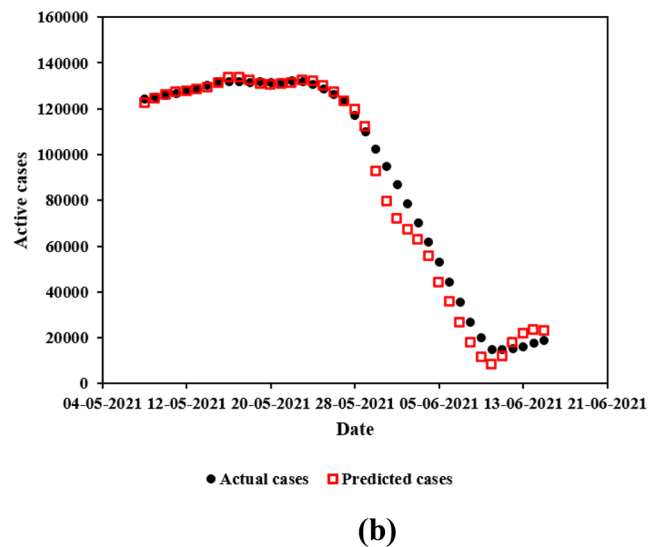
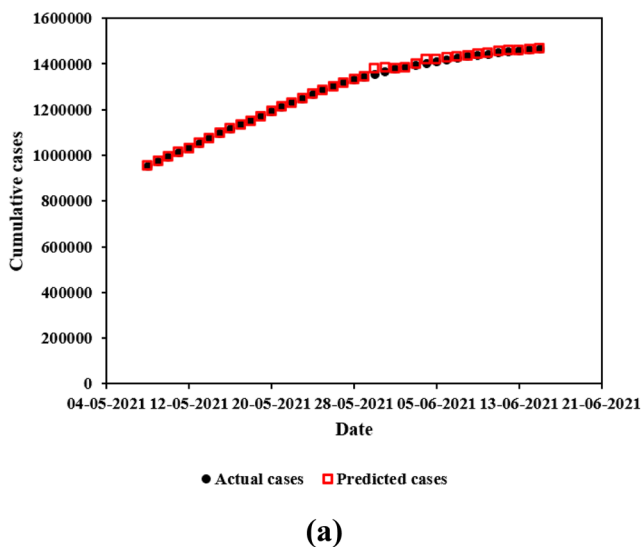


Fig. 11 Comparison of Predicted and Actual cases for **a** cumulative and **b** active cases during the periods of interventions

Conclusion

The daily variations of total number of cases and active cases have been fitted with cubic polynomial and quadratic polynomial respectively. The agreement between the predicted and actual data has been very good up to next 3 days and reasonable up to next 6 days. Considering the simplicity of the method and feasibility of its implementation with very modest computational resources and expertise, this can be very helpful for emergency planning of resources. One can come to the conclusion that cubic polynomial produces low APE and MAPE values for total cases, which, obviously, show a monotonic variation. For active cases, we employed quadratic and cubic polynomial functions. During prediction, situations are encountered where the active number of cases reaches some peaks. To capture the peak, dataset was shifted in the forward direction and the values are predicted for the next 6 days. It was found that the quadratic polynomial senses the presence of peak beforehand unlike the cubic polynomial and the quadratic polynomial yields less amount of error. Also, when there is a rise in the active number of cases, the quadratic polynomial proved to be a better tool in prediction than cubic polynomial. A comparison has been made for active cases between quadratic, cubic and quartic polynomial functions in terms of error values. The quadratic polynomial always produced better results than the other polynomial functions for active cases. Besides the prediction methods, a comparison has also been made regarding the predicted data range. One can predict values for $n + 1$ day, $n + 3$ days and $n + 6$ days, n being the last day of the dataset used for prediction. For each day, the errors or in other words the variations from the actual data have been calculated and compared. Expectedly, the error increases as one moves ahead in time with the minimum error seen for $(n + 1)$ th day and the maximum for $(n + 6)$ th day. APE values for $(n + 6)$ th day in case of prediction by quadratic polynomial are much less compared to the other functions. From this work, one can say that forecasting of active cases can be done using a quadratic polynomial, cumulative cases can be predicted by cubic polynomial and it will be a short-term forecasting for a week. Moreover, the predictions in the period of restricted movements show that the method of polynomials will work in challenging scenarios also. The authors have focussed on the active cases because the number of active cases will inform the administration about the necessary facilities needed to plan for immediate resource mobilization. Currently, people are experiencing the second wave of the pandemic and everyone must be prepared for any upcoming wave of this pandemic.

References

- Alvarez MM, González-González E, Trujillo-de Santiago G (2021) Modeling COVID-19 epidemics in an Excel spreadsheet to enable first-hand accurate predictions of the pandemic evolution in urban areas. *Sci Rep* 11:1–12. <https://doi.org/10.1038/s41598-021-83697-w>
- Batista M (2020) Estimation of the final size of the second phase of the coronavirus COVID 19 epidemic by the logistic model. medRxiv. <https://doi.org/10.1101/2020.03.11.20024901>
- Bell D, Nicoll A, Fukuda K et al (2006) Nonpharmaceutical interventions for pandemic influenza, international measures. *Emerg Infect Dis* 12:81–87. <https://doi.org/10.3201/eid1201.051370>
- Bhardwaj R (2020) A predictive model for the evolution of COVID-19. *Trans Indian Natl Acad Eng* 5:133–140. <https://doi.org/10.1007/s41403-020-00130-w>
- Botha AE, Dednam W (2020) A simple iterative map forecast of the COVID-19 pandemic. arXiv arXiv:1–9
- Chimmula VKR, Zhang L (2020) Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals*. <https://doi.org/10.1016/j.chaos.2020.109864>
- Fong SJ, Li G, Dey N et al (2020) Finding an accurate early forecasting model from small dataset: a case of 2019-nCoV novel coronavirus outbreak. arXiv 6:132–140. <https://doi.org/10.9781/ijimai.2020.02.002>
- Ganiny S, Nisar O (2021) Mathematical modeling and a month ahead forecast of the coronavirus disease 2019 (COVID-19) pandemic: an Indian scenario. *Model Earth Syst Environ* 7:29–40. <https://doi.org/10.1007/s40808-020-01080-6>
- Gao Y, Zhang Z, Yao W et al (2020) Forecasting the cumulative number of COVID-19 deaths in China: a Boltzmann function-based modeling study. *Infect Control Hosp Epidemiol* 41:841–843. <https://doi.org/10.1017/icc.2020.101>
- Giordano G, Blanchini F, Bruno R et al (2020) Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Springer
- Guleryuz D (2021) Forecasting outbreak of COVID-19 in Turkey; comparison of Box-Jenkins, Brown's exponential smoothing and long short-term memory models. *Proc Saf Environ Prot* 149:927–935. <https://doi.org/10.1016/j.psep.2021.03.032>
- Gupta R, Pal SK (2020) Trend analysis and forecasting of COVID-19 outbreak in India. medRxiv. <https://doi.org/10.1101/2020.03.26.20044511>
- Hu Z, Ge Q, Li S et al (2020) Evaluating the effect of public health intervention on the global-wide spread trajectory of Covid-19. medRxiv. <https://doi.org/10.1101/2020.03.11.20033639>
- Jia L, Li K, Jiang Y, et al (2020) Prediction and analysis of coronavirus disease 2019. arXiv
- Kapoor A, Ben X, Liu L, et al (2020) Examining COVID-19 forecasting using spatio-temporal graph neural networks. arXiv
- Komarova NL, Schang LM, Wodarz D (2020) Patterns of the COVID19 pandemic spread around the world: exponential vs power laws. medRxiv. <https://doi.org/10.1101/2020.03.30.20047274>
- Lana RM, Cruz OG, Villela D, et al (2020) Assessing the potential impacts of COVID-19 in Brasil : mobility, morbidity and impact to the health system. 1–14
- Magal P, Webb G (2020) Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy France and Germany. medRxiv. <https://doi.org/10.1101/2020.03.21.20040154>
- Okhueuse VA (2020) Mathematical predictions for COVID-19 as a global pandemic victor alexander okhueuse department of mathematics, faculty of natural and Applied sciences. Nasarawa State University

- Orihuel E, Sapena J, Navarro-Ortiz J (2021) An empirical algorithm for COVID-19 nowcasting and short-term forecast in Spain: a kinematic approach. *Appl Syst Innov* 4:1–18. <https://doi.org/10.3390/ASI4010002>
- Ranjan R (2020) Temporal dynamics of COVID-19 outbreak and future projections: a data-driven approach. *Trans Indian Natl Acad Eng* 5:109–115. <https://doi.org/10.1007/s41403-020-00112-y>
- Ribeiro MHD, da Silva RG, Mariani VC, dos Coelho L (2020) Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil. *Chaos Solitons Fractals*. <https://doi.org/10.1016/j.chaos.2020.109853>
- Rodgers JL, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. Published by : Taylor & Francis , Ltd . on behalf of the American Statistical Association Stable: <http://www.jstor.org/stable/2685263> 42:59–66
- Sameni R (2020) Mathematical modeling of epidemic diseases; a case study of the COVID-19 coronavirus. *arXiv* 1–19
- Sarkar K, Khajanchi S, Nieto JJ (2020) Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fractals* 139:110049. <https://doi.org/10.1016/j.chaos.2020.110049>
- Schlossmacher EJ (1973) An iterative technique for absolute deviations curve fitting. *J Am Stat Assoc* 68:857–859. <https://doi.org/10.1080/01621459.1973.10481436>
- Susser M, Adelstein A (1975) An introduction to the work of William Farr. *Am J Epidemiol* 101:469–476. <https://doi.org/10.1093/oxfordjournals.aje.a112117>
- Tomar A, Gupta N (2020) Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci Total Environ* 728:138762. <https://doi.org/10.1016/j.scitotenv.2020.138762>
- Verma MK, Asad A, Chatterjee S (2020) COVID-19 pandemic: power law spread and flattening of the curve. *Trans Indian Natl Acad Eng* 5:103–108. <https://doi.org/10.1007/s41403-020-00104-y>
- Wang S, Wang H, Zhang Y, Lu S (2020) Tracking and forecasting milestone moments of the epidemic in the early-outbreak: framework and applications to the COVID-19. *F1000Research* 9:1–21
- Wangping J, Ke H, Yang S et al (2020) Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared With Hunan, China. *Front Med* 7:1–7. <https://doi.org/10.3389/fmed.2020.00169>
- Weber A, Iannelli F, Gonçalves S (2020) Trend analysis of the COVID-19 pandemic in China and the rest of the world. *medRxiv*. <https://doi.org/10.1101/2020.03.19.20037192>
- Yusheng Z (2005) Time series forecasting competition among three sophisticated paradigms (<https://prsindia.org/covid-19/cases>)

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