Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A data driven epidemic model to analyse the lockdown effect and predict the course of COVID-19 progress in India

Bijay Kumar Sahoo a,b, Balvinder Kaur Sapra a,b,∗

a Radiological Physics and Advisory Division, Bhabha Atomic Research Centre, Mumbai, 400 094, India
b Homi Bhabha National Institute, Department of Atomic Energy, Mumbai, 400 094, India

A R T I C L E   I N F O

Article history:
Received 11 May 2020
Revised 15 June 2020
Accepted 18 June 2020
Available online 20 June 2020

Keywords:
COVID-19
Data driven model
Infected cases
Cross correlation
Time-lag analysis
Least square fitting
Mean recovery time
Prediction
Peak time
End time
Peak infected cases

A B S T R A C T

We propose a data driven epidemic model using the real data on the infection, recovery and death cases for the analysis of COVID-19 progression in India. The model assumes continuation of existing control measures such as lockdown and quarantines, the suspected and confirmed cases and does not consider the scenario of 2nd surge of the epidemic due to any reason. The model is arrived after least square fitting of epidemic behaviour model based on theoretical formulation to the real data of cumulative infection cases reported between 24 March 2020 and 30 May 2020. The predictive capability of the model has been validated with real data of infection cases reported during June 1–10, 2020. A detailed analysis of model predictions in terms of future trend of COVID-19 progress individually in 18 states of India and India as a whole has been attempted. Infection rate in India, as a whole, is continuously decreasing with time and has reached 3 times lower than the initial infection rate after 6 weeks of lock down suggesting the effectiveness of the lockdown in containing the epidemic. Results suggest that India, as a whole, could see the peak and end of the epidemic in the month of July 2020 and March 2021 respectively as per the current trend in the data. Active infected cases in India may touch 2 lakhs or slightly above at the peak time and total infected cases may reach over 19 lakhs as per current trend. State-wise results have been discussed in the manuscript. However, the prediction may deviate particularly for longer dates, as assumptions of model cannot be met always in a real scenario. In view of this, a real time application (COV-IND Predictor) has been developed which automatically syncs the latest data from the national COVID19 dash board on daily basis and updates the model input parameters and predictions instantaneously. This real time application can be accessed from the link: https://docs.google.com/spreadsheets/d/1fCwgnQ-dz4jOpLVMDHzcbEW14ZwOjIdEXm8TqfJDNVFA/edit?usp=sharing and can serve as a practical tool for policy makers to track peak time and maximum active infected cases based on latest trend in data for medical readiness and taking epidemic management decisions.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Corona virus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) [1,2]. The disease was first identified in December 2019 in Wuhan, the capital of China’s Hubei province. Since then, the numbers of cases have spread to all over the world. On March 11, 2020, the World Health Organization (WHO) formally declared the outbreak of novel corona virus as a Global Pandemic. As of June 01, 2020, a total of 6152,160 cases are confirmed in more than 227 countries and 26 cruise ships. There are 3142,964 active cases and 371,700 deaths [3].

The first case of the 2019–20 coronavirus pandemic in India was reported on January 30, 2020, originating from China and now India has become the largest affected country in Asia. As of June 01, 2020, the Ministry of Health and Family Welfare has confirmed a total of 190,535 cases, 91,819 recoveries and 5394 deaths in the country. For India, the fatality rate is relatively lower at 3.09%, against the global 6.63% as of 20 May 2020. Six cities account for about half of all reported cases in the country- Mumbai, Delhi, Ahmedabad, Chennai, Pune and Kolkata [4,5]. On March 22, 2020, India observed a 14-hour voluntary public curfew followed by a nationwide lockdown since March 24, 2020, besides several other measures such as quarantine of the suspected cases, public health guidelines on social distancing, frequent hand washing and wearing face mask to stem out of home for essential services.

Modelling and predicting the course of the outbreak in each region is important for the management and containment of the epi-
demic, and for balancing the impact from the public health vs. the economic crisis. Majority of COVID-19 epidemic models have originated from the SIR (Susceptible, Infected, and Recovered or Removed) model [6] and its many variations have been used in several countries, such as India [7], China [8,9], Italy [10,11] and Brazil [12]. These SIR-type models are useful for policy-decision makers to know the potential impact of pandemic and for prompting them to take early actions to minimise the impact. However, subsequent to breakout of pandemic, more information is required for a detailed planning, such as peak arrival of the pandemic, the number of hospital beds needed at the peak time, and taking decision on relaxing/lifting the lockdown, and finally returning to normal living. A recent study published in Nature [13] reveals that major non-pharmaceutical interventions and lockdown, in particular, have considerable effect on reducing transmission based on a large study carried out using real data in 11 European countries. Several studies are now becoming available on analysis of various epidemic control measures to contain the epidemic spread of COVID-19 in various countries [14–17].

In our study, we propose a data driven epidemic model to analyse the lockdown effect in India, 2nd largest population in world and to predict the course of COVID-19 progress for medical readiness using the latest data on cumulative infection cases and removed cases due to recovery and death. The model has the advantage that it does not depend on the susceptible population, a key parameter required for SIR type models. However, it has the disadvantage that it cannot be used when the epidemic has just started and the data are limited. The model has been implemented in a Google sheet for real time analysis of epidemic trend based on the latest data and it predicts various important parameters such as peak time, and number of active infected cases at peak time. This data will be helpful for arrangement of various medical recourses and taking epidemic related management decisions.

The manuscript presents the theoretical formulation and development of the model for cumulative, daily and active infected cases, application to COVID-19 scenario in 18 individual states of India and India, as a whole. It analyses the lockdown effect and predicts the course of epidemic progress as discussed in the following sections. Finally, a link to the real time application (COVIND Predictor) has been provided for daily updates and predictions.

2. Theoretical formulation and development of model

Let $N(t)$ is the number of total infected cases at time $t$. The rate of change of total infected cases can be expressed as

$$\frac{dN(t)}{dt} = \lambda_i(t)N(t) \quad (1)$$

where $\lambda_i(t)$ represents the infection rate at time $t$.

Generally, infection rate represents the number of contacts per person per unit time and it decreases with various control measures such as quarantine, lockdown etc. [18,19]. Let us consider a scenario of continued lockdown till the epidemic comes to a near end. Also, it is assumed that the infection rate in the population is highest at start of lockdown which decreases exponentially with increase in lockdown period and finally approaches zero after a sufficient time.

With this, the transient variation of infection rate subsequent to lockdown can be written as

$$\lambda_i(t) = \lambda_0 e^{-\frac{t}{\tau}} \quad (2)$$

where $\lambda_0$ is the initial infection rate at time of implementing lockdown. $\tau$ is the characteristic time of decrease which depends upon the societal factor, the extent of implementation of the lockdown in the society, number of quarantine person, number of samples tested etc.

Substituting the expression for $\lambda_i(t)$, Eq. (1) can be written as

$$\frac{dN(t)}{N(t)} = (\lambda_0 e^{-\frac{t}{\tau}})dt \quad (3)$$

If $N_0$ is the number of total infected cases at time of implementing lock down ($t = t_0$), the solution to the above equation can be written as

$$\log N(t) = \log N_0 + \lambda_0 \tau e^{-\frac{t_0}{\tau}} - \lambda_0 \tau e^{-\frac{t}{\tau}} \quad (4)$$

This can be re-expressed as

$$\log N(t) = k_1 - k_2 e^{-\frac{t}{\tau}} \quad (5)$$

where

$$k_1 = \log N_0 + \lambda_0 \tau e^{-\frac{t_0}{\tau}} \quad \text{and} \quad k_2 = \lambda_0 \tau \quad (6)$$

The number of infected cases, $N(t)$ (using Eqs. (5) and (6)) can be expressed as

$$N(t) = \frac{k_1 - k_2 e^{-\frac{t}{\tau}}}{k_2} \quad (7)$$

The model presented in Eq. (7) is also known as Gompertz function [20], initially proposed based on the nature of function expressive of the law of human mortality, with the assumption that mortality rate decreases exponentially as a person ages. A similar assumption has been made for the infection rate variation under lockdown scenario in the present study.

In real scenario, this particular assumption of the model cannot be met always. Hence there may be a change in the trend of infections rate due to various reasons (e.g. movement of migrant workers from one state to another, mild relaxation in lockdown rules). In such cases, it is advisable to do a trend analysis of latest data, update the model input parameters and make the predictions. However, the best approach is to develop a real time application which will fetch data from the national database dashboard at regular intervals and make predictions based on the latest trend analysis.

2.1. Daily new infected and removed cases

Differentiating Eq. (7) with respect to $t$, the number of new infected cases per day (i.e. daily new infected cases), $N_M(t)$ can be obtained as follows:

$$N_M(t) = \frac{dN(t)}{dt} = \frac{k_2}{\tau} e^{-\frac{t}{\tau}} \quad (8)$$

Since those who are admitted to the hospital either recover after a hospital stay of $T$ days, or may die after a similar number of days, there should be a delayed relationship between number of daily new infected cases, $N_M(t)$ and number of daily removed cases, $N_D(t)$ due to recovery and death.

Hence, $N(t)$ can be related to $N_M(t)$ by the following relation

$$N(t) = N_M(t - T), \quad ....... t > T \quad (9)$$

where $T$ is the mean recovery time of COVID-19 patients. This can be determined through time-lag correlation analysis between daily new cases and daily removed cases where epidemic has nearly come to an end.

2.2. Active infected cases

Similarly, the number of infected active cases, $N_a(t)$ can be estimated by taking the difference between cumulative new infected cases and cumulative removed cases up to time $t$, i.e.

$$N_a(t) = \int_{0}^{t} N_M(t')dt' - \int_{0}^{t} N_D(t')dt' = \int_{0}^{t} N_M(t')dt' - \int_{0}^{t} N_D(t' - T)dt' \quad (10)$$
Simplifying, we get

\[ N_0(t) = \frac{t}{t - T} N_0(t') dt' = \sum_{t' = t - T}^{t = t} N_0(t') \]  \hspace{1cm} (11)

This indicates that the number of active infected cases at time \( t \), is the sum of the number of daily new infected cases for a period of \( T \) preceding \( t \). The integration is used for the continuous functions of the model while the sum is used for the discrete real data.

2.3. Peak time of active infected cases

Maximum medical resources are required when the active cases attain maxima. Hence, predicting the maximum active cases and the time when this maximum will be attained is of utmost importance for planning and arrangement of medical resources such as number of hospital beds, ventilators, personal protective equipments for health care providers etc.

The rate of change of active infected cases at any time \( t \) can be written as

\[ \frac{dN_a(t)}{dt} = N_a(t) - N_r(t) \]  \hspace{1cm} (12)

Let \( t_p \) be the time at which the active cases attain peak; the occurrence of the peak is achieved if

\[ \frac{dN_a(t_p)}{dt} = 0 \]  \hspace{1cm} (13)

This condition leads to

\[ N_a(t_p) = N_r(t_p) \]  \hspace{1cm} (14)

After attaining peak, the newly recovered and deceased cases start to exceed the newly infected cases. The demand for medical resources, such as hospital beds, isolation wards and respirators, starts to decrease beyond this peak.

Using Eqs. (8) and (9), \( N_i(t) \) can be expressed as

\[ N_i(t) = \frac{k_2}{\tau} e^{\frac{-t}{\tau}} N(t - T) + \frac{k_2}{\tau} e^{\frac{-t}{\tau}} - k_1 e^{\frac{-t}{\tau}} \]  \hspace{1cm} (15)

Taking the ratio of \( N_0(t) \) and \( N_r(t) \) as given in Eqs. (8) and (15) respectively, and using condition given in Eq. (14) for the peak time, \( t_p \) can be obtained as:

\[ t_p = \tau \ln \left( \frac{k_2}{k_1} \frac{\tau}{T} e^{\frac{-t}{\tau}} - 1 \right) \]  \hspace{1cm} (16)

Eq. (16) can be used to obtain the time when active infected cases will attain peak with the information of characteristic time constant (\( \tau \)), recovery time (\( T \)) and fitting parameter \( k_2 \). Once the peak time \( t_p \) is estimated, the number of active infected cases at \( t = t_p \) can be estimated using Eq. (11).

3. Results and discussions

3.1. Estimation of model input parameters and validation

The first case of the 2019–20 corona virus pandemic in India was reported on January 30, 2020. India observed a nationwide lockdown since March 24, 2020 (55th day after 1st case) to control the epidemic in addition to several other measures. As of June 01, 2020, the Ministry of Health and Family Welfare has confirmed a total of 190,535 cases, 91,819 recoveries and 5394 deaths in the country. About 18 states have exceeded 100 confirmed cases as on May 01, 2020 [4,5]. These states have been considered for model analysis and prediction. State-wise break up of confirmed cases as on March 24 (start date of lockdown) and as on June 01, 2020 along with statistics of samples tested for these states is provided in Table 1.

The time series data of confirmed cases between March 24 and May 30, 2020 have been converted into logarithmic values as per requirement of the model (Eq. (5)) and least square fitted using ‘Origin’ software. The inbuilt function (ExpDec1) available in the software ‘Origin’ has been selected for fitting analysis which is expressed as:

\[ Y(x) = Y_0 + A_1 e^{-\frac{x}{\tau}} \]  \hspace{1cm} (17)

where \( x \) is taken as time (days) and \( Y \) is the logarithmic value of number of cumulative infected cases up to \( x \) days.

Comparing Eq. (17) with the model given in Eq. (5), the model parameters \( k_1, k_2 \) and \( \tau \) can be obtained from fitting parameters \((Y_0, A_1 \) and \( t_1)\) using the following equations

\[ k_1 = Y_0 \]  \hspace{1cm} (18)

\[ k_2 = -A_1 \]  \hspace{1cm} (19)

\[ \tau = t_1 \]  \hspace{1cm} (20)

The curve fitted to the data of India, as a whole, is shown in Fig. 1. Similar least square fitting exercise has been carried out for the selected states as well. The derived fitting parameters for the selected states and India, as a whole are also presented in Table 1.

Subsequently, models for each state and India as a whole have been tested against the real data of confirmed cases reported during June 01–10, 2020 to find the deviation of predicted values and testing the validity of the model. The maximum percentage deviation of model predictions has been given in Table 1. The results shows that deviations are within 10% in most of the states except a few states like Assam, Haryana, Karnataka, Odisha where it appears that a 2nd surge is emerging. In these states, model prediction may not be accurate with the existing fitting parameters and needs to be updated using the upcoming real data. It is important to note that the model is derived based on the certain assumptions as highlighted in model formulation section and does not consider scenario of 2nd surge due to any reason. Hence, model parameters needs to be updated in such changed scenario.

3.2. Effect of lock down on infection rate

Using Eq. (1), one can express the time-dependent infection rate as:

\[ \lambda(t) = \frac{dn(t)}{dt} = \frac{d}{dt} \left[ \log N(t) \right] \]

Using Eq. (2), the infection rate (per day) can be estimated by knowing the parameter \( \lambda_0 \) and \( \tau \). \( \tau \) is the characteristic time obtained directly from least square fitting analysis and is given in Table 1. \( \lambda_0 \) can be obtained as the ratio of fitting coefficient \( k_2 \) and \( \tau \).

Fig. 2 shows the plot of infection rate, predicted and that obtained from real daily and cumulative infected cases in India as a whole as a function of time. Time, \( t = 0 \) represents the start date of lockdown (i.e. 24 March 2020). As may be seen, there is a very good matching (<10% deviation) in the trend of predicted and real infection rate suggesting the validity of exponential model taken for infection rate. Least square fitting of the exponential model to the data points of real infection rate has yielded a \( R^2 \) value of 0.98 justifying the exponential behaviour model assumed for infection rate in a continued lockdown scenario. Initially the infection rate was around 0.15 per day which has come down to about 0.05 per day (about 3 times lower) after 6 weeks of lock down. The falling infection rate suggests the effectiveness of the lock down. If this trend continues, the predicted infection rate will reach one tenth of the initial infection rate (~0.015) by about 12 weeks from start date of the lockdown. However, India being a county with a huge population (about 135 Crore), even a very low infection rate can result in a large number of overall infected cases per day. If
the daily testing number is not adequate enough, it may be difficult to contain the spread of the epidemic in the community. Tracking the percentage of confirmed cases of the tested samples provides a better picture of the overall epidemic progress in the community. This parameter needs to be controlled by lowering the infection rate (i.e., number of contacts per person per unit time) by way of lockdown, masking of face and following social distancing norms as well as increasing the number of tests, detecting more and more infected cases and quarantining them as early as possible.

The characteristic time constant, \( \tau \), governs the decreasing trend of the infection rate. Higher this value, slower is the decrease in rate of infection. Table 1 provides the value of characteristic time constant for various states. Comparing the infection rate of various states and India as a whole with countries such as USA, Italy, Germany, and China, one can observe that the infection rate is quite low (\( \sim 2-3 \) times) in India. However, the decreasing trend is not as fast as USA, Italy, Germany and China [2]. This may be due to low testing of samples in India, particularly during the initial period of the epidemic.

### Table 1

State-wise cumulative confirmed cases as on start date of lockdown and as on June 01, 2020, tested sample statistics and derived model parameters through least square fitting to time series data of respective states.

| State /UT | Cumulative confirmed cases, \( N(t) \) as on March 24 (lockdown start) | Samples tested per 1 million people | Confirmed cases out of 100 samples | Derived input parameters of Model (Eq. (7)) | Characteristic time constant, \( \tau \) | Maximum% deviation from tested data (1–10 June) |
|-----------|---------------------------------------------------------------------|-----------------------------------|----------------------------------|-----------------------------------------|-----------------------------------------|---------------------------------------------|
| All states (India) | 571,90,648 | 7033 | 1 | 14.22±0.12 | 19.6±0.43 | 55.3±1.9 | 0.998 | 8 |
| Andhra Pradesh | 8 | 3571 | 140 | 10.1±0.4 | 27.9±7.8 | 44.1±6.0 | 0.981 | 9 |
| Assam | 0 | 3317 | 1.2 | 11.6±1.1 | 70.0±13.0 | 45.1±3.7 | 0.842 | 20 |
| Bihar | 3 | 3807 | 636 | 10.8±1.1 | 35.1±0.3 | 48.1±11.5 | 0.983 | 10 |
| Delhi | 30 | 19,844 | 11,821 | 12.9±0.3 | 30.2±1.0 | 53.9±4.2 | 0.989 | 10 |
| Gujarat | 34 | 16,794 | 3215 | 10.4±0.0 | 80.6±7.1 | 25.1±0.5 | 0.993 | 9 |
| Haryana | 30 | 2091 | 4375 | 10.8±0.2 | 99.2±15.8 | 34.1±6.8 | 0.881 | 19 |
| Jammu & Kashmir | 6 | 2446 | 12,218 | 9.6±0.3 | 80.0±9.7 | 32.0±4.6 | 0.991 | 8 |
| Jharkhand | 0 | 635 | 1781 | 8.9±0.2 | 40.2±19.1 | 45.1±1.3 | 0.973 | 10 |
| Karnataka | 41 | 3221 | 4448 | 10.1±1.2 | 35.1±7.2 | 43.1±4.1 | 0.891 | 20 |
| Madhya Pradesh | 7 | 8089 | 2046 | 9.5±0.1 | 115.3±12.7 | 22.1±0.8 | 0.982 | 10 |
| Maharashtra | 107 | 67,655 | 3828 | 12.8±0.1 | 28.1±0.8 | 53.1±1.3 | 0.998 | 6 |
| Odisha | 2 | 1948 | 3381 | 9.1±1.2 | 60.0±10.5 | 34.1±18.0 | 0.883 | 18 |
| Punjab | 29 | 2263 | 2928 | 8.1±0.5 | 370.0±18.3 | 16.0±6.0 | 0.960 | 14 |
| Rajasthan | 32 | 8831 | 5254 | 9.8±0.5 | 66.1±9.6 | 27.0±1.5 | 0.997 | 10 |
| Tamil Nadu | 18 | 22,333 | 6473 | 12.8±0.4 | 25.0±0.2 | 54.0±2.0 | 0.986 | 12 |
| Telangana | 37 | 2698 | - | 10.2±0.0 | 39.0±26.9 | 43.0±0.3 | 0.996 | 8 |
| Uttar Pradesh | 35 | 8075 | 1266 | 10.9±0.2 | 35.2±3.8 | 42.1±1.4 | 0.997 | 6 |
| West Bengal | 9 | 5501 | 2079 | 10.9±0.3 | 40.1±3.1 | 42.0±2.5 | 0.991 | 8 |

a) Model input parameters should be updated periodically based on trend in latest data to make the predictions more accurate.

![Fig. 1](image-url) Least square fitting of inbuilt exponential model 'origin' as given in Eq. (17) to the data of cumulative infected cases (\( N(t) \)) up to May 30, 2020 for India as a whole.
3.4. Time-lag correlation between daily new and removed cases and estimating mean recovery time

Now that the epidemic in Kerala appears to have come to an end, the data from this state has been used to perform cross correlation between daily new cases and removed cases due to recovery and death during the period March 14, 2020 and April 30, 2020. The plot of normalised correlation factor with respect to maximum value with different time lag is shown in Fig. 3. As may be seen, the correlation is found to attain maximum when the time lag between them is 15 days i.e. peak of daily new cases and daily removed cases is just lagged by 15 days. This is known as the mean recovery time, \( T \) for COVID patients. Typically for COVID-19 infection, the reported value of the mean recovery time varies from 14 to 16 days [2]. This recovery time of 15 days has been used for other states and India as a whole to estimate the peak of active infected cases.

3.5. Prediction of daily, cumulative and active infected cases

Subsequent to estimation of mean recovery time and model parameters through least square fitting exercise, predictions have been made for daily, cumulative and active infected cases with time. In this analysis, January 30, 2020 is designated as time, \( t = 1 \) (date of 1st reported case in India) and accordingly March 24, 2020
is $t = 55$ (start date of lock down). Predictions have been made only for $t>55$ till the cumulative infection cases attain saturation. Fig. 4 shows the plot of predicted mean, minimum, maximum of total infected cases, new daily cases and active infected cases in India with time. Figs. 5–7, show the plot of predicted total infected cases, new daily cases and active infected cases with time respectively in selected states of India. One can refer these plots to find the approximate time of peaking and near end of epidemic and number of active infected cases and saturation cases at peak and end time in various states of India and India as a whole.

3.6. Prediction of peak and end time of epidemic and maximum active infected cases

Table 2 provides state-wise results of predicted time to reach peak of the epidemic, time to attain 99% of the total infected cases (= end time of epidemic) and number of active and total saturation of infected cases with lower and upper bound value considering the error margin in the derived model input parameters. These results are very much useful for planning and arrangement of medical resources.

Fig. 4. Predicted mean, minimum, maximum of total infected cases, new daily cases and active infected cases in India as a whole based on the data trend analysis up to May 30,2020.

Fig. 5. (a) and (b): Predicted total infected cases in various selected states of India based on the data trend analysis up to May 30,2020.
Results suggest that India, as a whole, could see the peak of the epidemic in the month of July 2020. Some of the states such as Madhya Pradesh, Punjab have already seen the peak of epidemic by this time while Andhra Pradesh, Assam, Gujarat, Jammu & Kashmir, Jharkhand, Karnataka, Rajasthan may see the peak in the month of June 2020. Maharashtra, Haryana, Odisha, Telangana, Uttar Pradesh, West Bengal may see the peak of epidemic in the month of July 2020 while Delhi, Tamil Nadu and Bihar may see the peak in peak of the epidemic in August 2020.

Results on active infected cases at peak time suggest that active infected cases for India, as a whole, may go a little above 200 K (K stands for thousands, here onwards). The most affected states- Maharashtra, Tamil Nadu and national capital Delhi may see the active infected cases up to 57 K, 42 K and 41 K respectively. Gujarat, West Bengal, Uttar Pradesh and Rajasthan may see the number of active infected cases between 5 K and 10 K at the peak time while remaining states may see the number of active infected cases below 5 K.

The time to reach 99% of the total expected infected cases is considered as the end of the epidemic by which most of the active infected cases have recovered. Results suggest that India, as a whole, could see the end of the epidemic in the month of March 2021 with total cases over 19 Lakhs. States like Bihar, Odisha, Karnataka, and Haryana could see end of the epidemic by December 2020 while most of states show end of epidemic by March 2021. The most affected state Maharashtra may see the end of the epidemic by February 2021 with total cases of about 5 Lakhs. The predictions emphasise the need to follow all precautionary measures
Table 2
State-wise predicted peak time, time to attain 99% of the total expected cases and number of active and total saturation infected cases based on the data trend analysis up to May 30, 2020.

| State/LT | Time to reach peak of the active infected case<sup>a</sup> | Number of active infected cases at peak time<sup>a</sup> | Time to reach 99% of total infected cases<sup>a</sup> | Number of total expected infected cases<sup>a</sup> |
|----------|---------------------------------------------------|--------------------------------------------------|-----------------------------------------------|--------------------------------------------------|
|          | Most likely | Lower Bound | Upper Bound | Most likely | Lower Bound | Upper Bound |
| India as a whole | July 2020 | 2,07,437 | 1,82,543 | 2,23,235 | March 2021 | 21,00,305 | 18,86,453 | 24,25,385 |
| Andhra Pradesh | June 2020 | 3443 | 3010 | 3820 | March 2021 | 47,290 | 42,657 | 52,964 |
| Assam | June 2020 | 2836 | 2057 | 3759 | Dec 2020 | 10,439 | 7523 | 14,547 |
| Bihar | Aug 2020 | 4476 | 3845 | 5648 | Nov 2020 | 37,232 | 34,623 | 41,347 |
| Delhi | Aug 2020 | 41,832 | 37,457 | 45,854 | March 2021 | 4,82,346 | 4,00,906 | 5,65,657 |
| Gujarat | June 2020 | 7790 | 6897 | 8229 | March 2021 | 65,145 | 60,345 | 70,127 |
| Haryana | July 2020 | 6567 | 5078 | 7923 | Dec 2020 | 54,567 | 46,987 | 67,975 |
| Jammu & Kashmir | June 2020 | 2819 | 2532 | 3118 | Jan 2021 | 23,891 | 18,302 | 28,017 |
| Jharkhand | June 2020 | 1490 | 1120 | 1812 | March 2021 | 14,532 | 10,157 | 19,124 |
| Karnataka | June 2020 | 3678 | 3218 | 4057 | Dec 2020 | 28,216 | 20,622 | 34,213 |
| Madhya Pradesh | May 2020 | 3497 | 3147 | 3898 | March 2021 | 30,894 | 25,478 | 35,568 |
| Maharashtra | July 2020 | 57,350 | 53,459 | 61,389 | Feb 2021 | 5,03,684 | 4,10,786 | 6,02,932 |
| Odisha | July 2020 | 2136 | 1534 | 2966 | Dec 2020 | 28,051 | 21,012 | 34,112 |
| Punjab | May 2020 | 1565 | 1263 | 1760 | March 2021 | 21,017 | 16,125 | 26,432 |
| Rajasthan | June 2020 | 4381 | 3876 | 4912 | March 2021 | 35,358 | 30,851 | 40,770 |
| Tamil Nadu | Aug 2020 | 42,886 | 39,093 | 45,236 | March 2021 | 47,236 | 39,210 | 50,436 |
| Telangana | July 2020 | 3439 | 3012 | 3912 | March 2021 | 47,689 | 42,814 | 52,121 |
| Uttar Pradesh | July 2020 | 8188 | 7497 | 8958 | March 2021 | 94,740 | 87,459 | 99,702 |
| West Bengal | July 2020 | 9838 | 8378 | 10,971 | March 2021 | 1,00,065 | 85,849 | 1,20,578 |

<sup>a</sup> Results may deviate depending upon the change in the trend of the latest data. It is advisable to check the latest predictions by real time model using the link: https://docs.google.com/spreadsheets/d/1fcwgnQ-dz4j0YWVHDHUbEcEW1423wOjJdExm8TjlJDWNA/edit?usp=sharing.

As it is well known, model assumptions cannot be kept always in real scenarios and hence the prediction may deviate depending upon change in the trend of latest data due to several factors such as movement of migrant workers and relaxation in epidemic control measures. In this context, a real-time application (COV-IND Predictor) has been developed by implementing the above model in a Google sheet which automatically syncs the latest data from national COVID19 dash board [5] on daily basis and updates the model input parameters using the inbuilt forecast function and makes the predictions instantaneously for future dates. The application can be accessed from the link: https://docs.google.com/spreadsheets/d/1fcwgnQ-dz4j0YWVHDHUbEcEW1423wOjJdExm8TjlJDWNA/edit?usp=sharing.

While the data predictions in this manuscript are based on real data only up to May 30, 2020, it is advisable to check the latest predictions from the above link which gives a more reliable prediction of the COVID-19 scenario in India, as a whole, and in individual selected state based on the latest trend in the COVID-19 infected cases. This real-time application can serve as a helping hand for policy makers to track peak time and maximum active infected cases based on latest trend in data for medical readiness and for taking epidemic management decisions periodically.

4. Conclusion

We propose a data-driven model to track and predict the course of the epidemic. Many parameters characterizing an epidemic can be determined from the model using the available latest data which can be validated by a few real data sets. Subsequently, the model can be used for predictions. This presented approach can be applied not just to the current Covid-19 epidemic, but also, in general, to future epidemics. The model gives best predictions with online type predictor, utilising latest data to update the model input parameters periodically and predict the course of epidemic for the next two weeks. The model is of special significance for predicting the approximate peak time and end time of the epidemic so as to keep a readiness for maximum resources during the peak time. The model is able to well capture the observed decrease in the infection rate post lockdown, thus confirming the effectiveness of lockdown in containing the epidemic. The model has been implemented in a Google sheet which can serve as a practical tool for epidemic management decisions such as medical resource planning, required number of daily testing and ultimately relaxing lockdown rule and regulation in order to balance the impact from the public health vs. the economic crisis [21–24].

Declaration of Competing Interest

The authors declare no competitive interests.

CRediT authorship contribution statement

Bijay Kumar Sahoo: Methodology, Formal analysis, Writing - original draft. Balvinder Kaur Sara: Conceptualization, Formal analysis, Visualization, Writing - review & editing.

Acknowledgement

Authors would like to thank Shri Suresh Babu RM, Associate Director, Health, Safety and Environment Group of BARC for his valuable suggestions and inputs during internal review of the manuscript.

References

[1] Coronavirus disease 2019 (COVID-19)--Symptoms and causes. Mayo Clinic; https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/yyc.201909063 [Retrieved 1 May 2020].
[2] Huang N.E., FangliQiao K.-K.T.. A data driven model for predicting the course of COVID-19 epidemic with application for China, Korea, Italy, Germany, Spain, UK and USA. medRxiv 2020; 10.1011/2020.03.28.2004677.
[3] COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). ArcGIS, Johns Hopkins Universit; 2020 Retrieved 1 June. https://github.com/CSSEGISandData/COVID-19.

[4] Home, Ministry of Health and Family Welfare, Government of India. www.mohfw.gov.in [Retrieved 1 June 2020].

[5] COVID-19 INDIA CORONAVIRUS PANDEMIC. https://www.covid19india.org/ [Retrieved 1 June 2020].

[6] Kermack W, McKendrick A. A contribution to the mathematical theory of epidemics. Proc R Soc Lond A 1927;115:700–21.

[7] Singh R, Adhikari R. Age-structured impact of social distancing on the COVID-19 epidemic in India.arXiv preprint arXiv 2020;2003.12055.

[8] Lin Q, Zhao S, Gao D, Lou Y, Yang S, Salihu S, Musa M, Wang H, Cai Y, Wang W, Yang L, He D. A conceptual model for the corona virus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int J Infect Dis 2020;93:211–16.

[9] Khrapov P, Loginova A. Mathematical modelling of the dynamics of the Corona virus COVID-19 epidemic development in China. Int J Open Inf Technol 2020;8(4):13–16.

[10] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solutions Fractals 2020;134:109761.

[11] Caccavo D. Chinese &Italian COVID-19 outbreaks can be correctly described by a modified SIRD model 2020; medRxiv: 10.1101/2020.03.19.20039388

[12] Bastos S.R, Cajeiro D.O. Modelling and forecasting the Covid-19 pandemic in Brazil. arXiv preprint arXiv 2020;2003.14288

[13] Imperial College COVID-19 Response Team Flaxman S, Mishra S, Gandy AH, Juliette T, Thomas AM, Helen C, Charles W, Harrison Z, Tresnia B, Jeffrey WE, Mélodie M, Azra G, Christl AD, Steven MR, Michaela ACV, Neil MF, Iucy CO, Samir B. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020. doi:10.1038/s41586-020-2405-7.

[14] Juliet B, Delia E, Johan G, David LH, Chikwe I, Gary K, H CL, Ziad M, Myoung-don O, Amadou AS, Anne S, Kumnuan U, Lotha HW. COVID-19: towards controlling of a pandemic. Lancet 2020;395(10229):1015–18.

[15] Yonghong X, Mali ET. Taking the right measures to control COVID-19. Lancet Infect Dis 2020;20(5):523–4.

[16] Di W, Jianyun L, Yanhui L, Zhoubin Z, Lei L. Positive effects of COVID-19 control measures on influenza prevention. Int J Infect Dis 2020;95:345–6.

[17] Balaji KM, Rana S. COVID-19 in INDIA: strategies to combat from combination threat of life and livelihood. J Microbiol Immunol Infect 2020;53(3):389–91.

[18] Biao T, Nicola LB, Qian L, Sanyi T, Yanni X, Jianhong W. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV). Infect Dis Model 2020;5:248–55.

[19] Ugo Avila-Ponce de L, Angel G., Perez C., Eric Avila-Vales. A data driven analysis and forecast of an SEIARD epidemic model for COVID-19 in Mexico. 2020; https://www.researchgate.net/publication/340776717

[20] Gompertz B. On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies. Philos Trans R Soc Lond 1825;115:513–85.

[21] Peak CM, Childs LM, Grad YH, Buckee CO. Comparing non-pharmaceutical interventions for containing emerging epidemics. Proc Natl Acad Sci 2017;114(15):4023–8. doi:10.1073/pnas.1616438114.

[22] Dhillon RS, Srikrishna D. When is contact tracing not enough to stop an outbreak? Lancet Inf Dis 2018;18:1302–4. doi:10.1016/S1473-3099(18)30656-X.

[23] Pang X, Zhu Z, Xu F, Guo J, Gong X, Liu D, Liu Z, Chen DP, Feikin DR. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing. JAMA 2003;290(24):3215–21.

[24] Wang G, Huang NE, Qiao F. Quantitative evaluation on control measures for an epidemic: a case study of COVID-19. Sci Bull 2020;65. doi:10.1360/ TB-2020-0159.