NEWS AND NOTES

Understanding Trend of the Covid-19 Fatalities in India – V.P. Dimri^{1*}, Shib S. Ganguli¹, R.P. Srivastava^{2†}. ¹CSIR-National Geophysical Research Institute, Hyderabad-500007, India. ²Equinor ASA, Sandslivegen 90, Bergen, 5254, Norway, formerly at CSIR-NGRI, (E: vpdimri@gmail.com*, ravi.ngri@gmail.com[†])

Introduction

The number of COVID-19 cases and related deaths are increasing at an alarming rate in several countries, and some countries have experienced mild spread. Globally, COVID-19 has caused an estimated 4,103,241 cases and 2,82,728 deaths until now (11/05/2020). Wuhan city in Hubei province of China is deemed as its epicenter, where the first COVID-19 case was reported in December 2019. Since, then the virus has spread over 215 countries and territories around the world, with varying spread rates and fatalities (Fanelli and Piazza, 2020). The worldwide lockdown and social distancing have played a great role to keep the reproduction number in check, and thus controlling the outbreak (Kupferschmidt, 2020). So far reported fatality rate of the corona virus (COVID-19) pandemic is somewhere between normal flu (0.1 to 0.2 %) to severe acute respiratory syndrome (SARS) family of viruses (8 to 10 %). In India, so far the fatality rate is about 3 %, which is half of the world average.

Forecasting the future of pandemics, including COVID-19, is difficult. The accuracy of an epidemic and pandemic forecasting largely depends on the data availability and uncertainty estimation. Despite this fact, still forecasting utilizing mathematical models has become crucial to better understand the present circumstances and plan the future. Mathematical models are invaluable tools for analyzing the spread and control of infectious diseases (Hethcote, 2000). In the present work, we provide a statistical forecasting approach for the COVID-19 cases in India and its potential implications for planning and decision making to contain the same.

Epidemic Breakout Simulation Model

The standard epidemiological modelling approach usually follows the reproduction number of the diseases, and models it using a well-known system of differential equations known as SIR (Susceptible - Infectious - Recovered) model (Batista, 2020). The differential equations for a generalized SIR model can be written as follows:

$$\frac{dS(t)}{dt} = \beta \frac{S(t) I(t)}{N} \quad (1)$$

$$\frac{dI(t)}{dt} = \beta \frac{S(t) I(t)}{N} - \gamma I(t) \quad (2)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (3)$$

where $S(t)$ = number of susceptible cases at time 't'; $I(t)$ = number of Infected persons and $R(t)$ = number of recovered cases at time 't', respectively. β is the contact rate and $1/\gamma$ is the average period of infections. The constant $N = S + I + R$ is the fixed number of populations in a certain region/country. In this case, the births and natural death are not modelled. The initial conditions can be expressed as $S(0) = S_0$, $I(0) = I_0$ and $R(0) = R_0$. Now, the parameters β , γ , and the initial values S_0 and I_0 (as obtained from available data) are required to utilize the

SIR model. The required parameters can be assessed by minimizing the difference between the actual and predicted number of cases.

Moreover, there are several papers available on the SIR model and generalized SEIR model (please refer to Hethcote, 2000; Fanelli and Piazza, 2020; Peng et al. 2020), hence we are not discussing it more here since it is not our main focus. The objective of listing equations and variables within the SIR model is to give an idea to a general reader about the various considerations involved in the model.

Often statistical epidemiological analysis applied in such cases assumes an exponential outbreak, which is to some extent true in a very early stage of the epidemic. The exponential model is based on the fixed reproduction number of the disease. The total number of cases can be obtained by considering the fact that each infected person contaminates the other susceptible one (where $n > 1$ known as reproduction number):

$$n^{(t/T)} = e^{at} \quad (4)$$

where, t and T represents time in days and incubation period, which relies upon the type of disease; a is coefficient of exponential series. In the case of COVID-19, the incubation period ranges somewhere between 10 to 14 days. It is to note that the above equation does not assume any kind of inhibition or intervention.

Analysis of COVID-19 Pandemic in Indian Scenario

In order to forecast the COVID-19 pandemic behavior in India, we first examined the total number of confirmed cases and the rate of daily increase in the number of cases reported in India (Fig. 1a and b). Besides, Fig. 2 (a and b) depicts the total confirmed deaths and the rate of daily increase in the number of deaths. The data is fetched from the web-page maintained by Johns Hopkins University (https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/). This is a dynamic data, and the figures below show data starting from 22 January 2020 until 10 May 2020. Please note that the number of confirmed cases is dependent on how much testing has been done, hence we choose to further analyze reported number of COVID-19 death cases, which are much more reliable numbers, than the infected cases.

We looked at the total death data in India and plotted it on a log-log scale (Fig. 3). From equation (4) it is obvious, if the number of deaths follows exponential increase, then in the log-normal plot they will appear linearly increasing, whereas if they follow power-law increase then they will appear linear in log-log scale (Dimri, 2005; Dimri et al. 2012). It is obvious from Fig.3, that deaths until April 12, 2020 follow power-law and the predicted deaths are in good agreement to the actual. However, this relation (data trend) appears to be invalid when the latest death data is analyzed hence, we did not use the power-law relation for further analysis. In order to understand and forecast COVID-19 behavior until June 10, 2020, we attempted a quadratic and exponential fit to the data. As evident from Fig.4, the total number of deaths after one month from now (10/05/2020) would reach approximately 6182, if the quadratic trend is followed. Otherwise, the

Fig.1. (a) Total number of reported confirmed COVID-19 cases in India and (b) Reported increase in number of cases per day as on 10 May 2020.

Fig.2. (a) Total number of COVID-19 related deaths and (b) increase in deaths per day in India as on 10 May 2020.

death toll can reach as high as about 21,628 ($\pm 5\%$) in the extreme case, as indicated by the exponential trend. Table 1 represents the week wise death values until 10 June, 2020, as forecasted using quadratic simulation. Note that these predictions are statistical, and

assume no interventions which are in place. Results are based on the most feasible scenarios and may vary in actual cases, depending upon the situation.

Our two approaches, one based on exponential fit to the data, and another based on simulating quadratic time series, gave completely different predictions, however, we expect them to behave like this, because they are fundamentally different. We believe that these two methods represent upper and lower bound scenarios of possible COVID-19 deaths in India, which could be a good suggestion to policy makers and authorities. None of these methods are meant to predict the epidemic and determine the life span of a pandemic in terms of the number of days/ months. The methods presented here are very good indicators to model dynamic data and draw short term

Fig.3. Shows a log-log plot of the total number of COVID-19 related deaths in India (blue curve), starting from 11 March 2020 until 12 April 2020. A power law curve fit to the data (excluding initial 7 data points to obtain a more robust fit) is shown in orange color. According to power law, two prediction dates marked by red plus sign correspond to 19 and 26 April 2020 respectively, and the corresponding death toll was approximately 509 and 909.

Table 1. Week wise predicted number of deaths together with minimum and maximum values due to COVID-19 pandemic in India, based on quadratic simulation model.

Date	Number of deaths		
	Min	Expected	Max
May 13, 2020	2220	2399	2579
May 20, 2020	2975	3180	3385
May 27, 2020	3824	4070	4317
June 03, 2020	4768	5071	5375
June 10, 2020	5807	6182	6557

Fig. 4. Shows total COVID-19 deaths in India and simulation of COVID-19 deaths with confidence intervals of 95% (red curve). Please note this is not just a curve-fitting exercise (actual reported deaths are represented by blue curve), rather a pure simulation of time series.

inference. For instance, in Fig.4 if the actual number of deaths go below quadratic fit, then we can assume that the outbreak is in the decline phase, however, if the number of deaths increases more than the fit, then we cannot rule out the possibility of further outbreak. Results presented in Fig. 4 are more realistic and fundamentally more robust, where the effects of constraints such as social distancing and lockdown are baked within the data used to create a simulated time series.

In order to forecast the final size of the COVID-19 pandemic in India, a data-driven SIR model is employed. The COVID-19 pandemic life cycle (spread) pattern is anticipated to be “bell-shaped”, as indicated by Fig. 5 (b). This helps to detect the inflection point and the peak of infections to distinguish acceleration and deacceleration phases. The results suggest that infected cases in India are at the acceleration phase (initiation of its inflection point) and the size of the epidemic will be about 167,200 infected cases until 10 June 2020 (Fig.5a). Note that the early segment of the curve is fitted with data, whereas the remaining segment is predicted based on the SIR model. This model assumes that it is a rational portrayal of the one-stage epidemic and represents the dynamic process of COVID-19 infections in a population over a specific time. Hence, the forecast is as good as data are. The forecasting may vary with new data or altered data (Batista, 2020).

CONCLUSIONS

In this note, we have presented the analysis of short-term forecasting of COVID-19 infections in India based on SIR, quadratic and exponential approaches. If the situation remains stable and data collection methods do not change much, then the predicted total number of deaths due to the pandemic until 10 June 2020 will be about 6182. These two predictions provide us bounds of deaths forecasted for two different scenarios, i.e. usual trend and extreme

Fig.5. Shows results from SIR model: (a) total number of predicted together with observed cases, and (b) number of COVID-19 cases per day in India. The different colour shades represent various epidemic phases, e.g. red=fast growing phase; yellow=transition to steady-state phase, and green=conclusion phase (plateau stage).

case. Uncertainty prevails in the future, regardless of model and data, which need to be kept in mind while doing or reading any prediction, and forecasting COVID-19 is not an exception. It is commendable that India despite having the largest population density managed to limit the outbreak with its timely intervention in terms of country lockdown. Nevertheless, the present analysis would be very helpful to deal with the unwanted future rising trend of the COVID-19 infections in India and develop guidelines or strategies to contain the pandemic.

Acknowledgement: We thank Director, CSIR-NGRI for support and permission to publish the work.

References

- Batista, M. (2020) Estimation of the final size of the coronavirus epidemic by the SIR model. University of Ljubljana, Slovenia.
- Dimri, V.P., Srivastava, R.P. and Vedanti, N. (2012) Fractal Models in Exploration Geophysics: Applications to Hydrocarbon Reservoirs. Elsevier Science, Vol. 41, p.165.
- Dimri, V.P. (2005) Fractal Behaviour of the Earth System. Springer, Berlin Heidelberg, p. 208.
- Fanelli, D. and Piazza, F. (2020) Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons and Fractals, v.134, pp.109761. <https://doi.org/10.1016/j.chaos.2020.109761>.
- Kupferschmidt, K. (2020) Ending coronavirus lockdowns will be a dangerous process of trial and error. Science, DOI:10.1126/science.abc2507.
- Hethcote, H.W. (2000) The Mathematics of Infectious Diseases. SIAM Review, v.42(4), pp. 599-653.
- Peng, L., Yang, W., Zhang, D., Zhuge, C., Hong, L. (2020) Epidemic analysis of COVID-19 in China by dynamical modeling. medRxiv. DOI: <https://arxiv.org/pdf/2002.06563.pdf>.

DOI: 10.1007/s12594-020-1492-2