Modeling the COVID-19 outbreak in India — with an analysis of the socio-economic scenario

Debashis Saikia · Kalpana Bora · Madhurjya P. Bora

Abstract We present a modeling and analysis of the COVID-19 outbreak in India with an emphasis on the socio-economic composition, based on the progress of the pandemic in 11 federal states where the outbreak is the largest in terms of total number of infectives. Our model is based on the susceptible-exposed-infectives-removed (SEIR) model, including an asymptomatic transmission, time dependent incubation period and time dependent transmission rate. We carry out the analysis with the available disease data up to the end of August 2020, with a projection of 54 days into the months of September and October 2020, based on the past data. Overall, we have presented a projection up to 400 days (till April 18, 2021) for India. We also find the existence of a critical day, signifying a sudden shift in the transmission pattern of the disease, with interesting relation of the behavior of the pandemic with demographic and socio-economic parameters. The results of this work can be used as a future guidance to follow in case of similar pandemics in the developing countries.

1 Introduction

The outbreak of SARS-CoV-2 disease has been one of the worst pandemics the world is experiencing in the present century. The disease had started at Wuhan, China in December 2019 and has currently infected almost all the countries across the globe. WHO declared this new coronavirus disease (also known as COVID-19) as a pandemic on 11 March, 2020 (33,34). As there is no vaccination for COVID-19, mathematical models help us analyse the evolution of the disease and also helps us predict the future scenario of the disease transmission, based on which governments may take necessary actions to suppress future transmissions. Many mathematical models are being formulated to study the spread of COVID-19 in different countries and to predict the evolution of the outbreak (24,20,39,13,19,22,17,29,36,26,34). As of mid-January 2021, total global cases of the pandemic has reached more than 90 million (8).

In India, the first COVID-19 case was confirmed on 30th of January, 2020 (7). In the early phase of the epidemic in India, certain events during March, 2020 in the states of Delhi and Punjab have appeared to become the ‘super-spreaders’ of COVID-19 and Delhi became the hotspot in India (33,27) along with the city of Mumbai. One unique cause which fuelled the spreading of the disease in certain states such as in Bihar, Assam, West Bengal, and Odisha was the movement of migrant workers during May 2020 (31,37).

The Indian scenario as far as the pandemic is concerned is quite intriguing. Some of the major facts which justify an India-specific modeling are the huge population size and its density, the status of the economy, extreme diversity in the

Debashis Saikia
Physics Department, Gauhati University, Guwahati 781014, India

Kalpana Bora
Physics Department, Gauhati University, Guwahati 781014, India

Madhurjya P. Bora (Corresponding author)
Physics Department, Gauhati University, Guwahati 781014, India
E-mail: mpbora@gauhati.ac.in
working population and economic status of the people, diversity in culture, language, and the way of living. The socio-economic perspective of the response to the pandemic and its related measures assume immense importance when one realises the fact that the social behavior of the population can vary widely from state to state as the lifestyle and culture, which govern the social behavior are very different as one travels across the length and breadth of the country. The same can be said for economy. The gross domestic product (GDP) stands at a staggering 450 billion USD for the state of Maharashtra to minimal 1.2 billion USD for the union territory of Andaman and Nicobar Islands (5, 10). There are more than 19,500 languages and dialects are spoken in India as mother tongues of which about 120 languages are spoken by more than 10,000 people in India (1). It is also important to note that the lifestyles of the people of one state is greatly influenced by its culture, which is identified with the language spoken. The food habits of the people also vary widely from state to state. For example, meat and poultry products are quite common in the eastern part of the country while vegetarianism is very popular in some northern states like Haryana, Rajasthan etc. According to government surveys (4), about 30% of the Indian population are vegetarian. As these socio-economic behavior is closely related to the progress of the pandemic, the response of the state population to pandemic is worth studying. The population of many Indian states is comparable to some countries. The Table 1 shows how the populations of certain Indian states can be compared with some countries of the world (6).

Table 1  Comparison of population of some of the Indian states (6).

| Sl No | Indian States       | Countries                      |
|-------|---------------------|--------------------------------|
| 1.    | Bihar, Maharashtra | Mexico, Japan                  |
| 2.    | West Bengal         | Egypt                          |
| 3.    | Rajasthan           | Iran, Germany                  |
| 4.    | Karnataka           | France, UK                     |
| 5.    | Gujarat             | Italy                          |
| 6.    | Odisha              | Spain, Argentina, Ukraine      |
| 7.    | Assam               | Saudi Arabia, Canada, Morocco, Poland |
| 8.    | Delhi               | Chile, Netherlands, Romania, Ecuador, Kazakhstan, Guatemala |

In this work, we have constructed a compartmental non-autonomous disease model based on the well known susceptible-exposed-infectives-removed (SEIR) model (23, 27). We have applied this model to eleven Indian states, which have the highest number of infectives of COVID-19, except for the states of Andhra Pradesh and Telangana, which were newly created states and do not have the updated census data. We note that, very early into the pandemic, India has witnessed a prolonged national level lockdown followed by a huge migration, primarily affecting the people with low income and employed in the informal sector. Besides there were certain large scale localised events at the time when the lockdown was being imposed (33, 21). Both these factors are believed to have affected the way the pandemic has progressed in India. Our analysis of the federal states, which is based on this model, suggests certain uniform and unique behavior the way the pandemic has advanced. We have limited our analysis to the end of August 2020, as most of the establishments and business have started to open, including some educational institutions, from September onwards. Besides, from September onwards, the winter festival season begins in India during which the infections are expected to rise considerably. In what follows, we first describe and develop our model. Next, we present an equilibria analysis for our model and calculated the basic reproduction number $R_0$. Then we analyse the progress of the pandemic for eleven Indian states, and the pandemic situation for the whole country and provide our projections. Finally we summarise and conclude with the novel ideas used and main results obtained in the work.
2 The model

2.1 The limitations on data

It is worth noting that though the gross official COVID-19 data is maintained by the Ministry of Health and Family Welfare (MoHFW) in India \(^9\), it is difficult to obtain the fine scale reliable data. For example, while we know the daily statistics of the newly infected and recovered people as well as the daily deaths, the reliable data about patients in non-critical and critical stage (patients requiring ventilators), hospital discharge statistics with reference to different age group, statistics about patients whether they are home-quarantined or officially quarantined (the so called COVID Centres) etc. are not easily available. Based on these observations, we have modeled the pandemic on the basis of what data is available in the public domain. For modeling the state data, we have use the crowdsourced data, which is maintained by https://www.covid19india.org \(^7\). It should be noted that this dataset closely match the official data provided by the MoHFW and the data maintained by John Hopkins’ COVID database \(^8\).

2.2 The facts and assumptions

It is understood that a considerable part of the infectious population is asymptomatic. It is important to note that our asymptomatic population is not what is reported after detection of the disease in an asymptomatic person. When an asymptomatic person is listed as an infected person, it adds to the total number of infectives, irrespective of whether the disease is asymptomatic or symptomatic. However, there is a large population of the people, which are potential infectives but never detected. This population is always hidden as an asymptomatic person does not show any kinds of signs of the disease but equally infectious just like as a person who is known to have the disease. At present, there is no known method to isolate this population except large scale systematic testing of the general population.

We also know that the disease has an incubation period after a person being exposed to the virus. However, the incubation period is only known for the population who is known to have been exposed. The incubation stage for the asymptomatic population simply does not exist. We further assume that the number of death due to causes other than COVID-19 is negligible during the pandemic period. We also assume that once a person is known to have been exposed to the virus, the person is being quarantined or being closely monitored, thereby preventing such a person to become asymptomatic carrier.

As now almost all nations have resorted to a lockdown or other restrictive measures once they have detected that there is a section of population known to have contracted the disease, the infection rate definitely decreases over time. Besides, during the initial phase of the spread of the disease, it is expected that the infection rate through the asymptomatic population remains negligible only to become significant as the disease spreads over time. The other significant deviation from the classic models is the inclusion of time dependence of rate at which the exposed population starts contracting the disease. We assume that at the start of the pandemic, the incubation period is negligible as most of the population starts in the infected compartment. Based on these observations, we can now write down the following non-autonomous disease model which is sourced from the basic susceptible-exposed-infectives-removed (SEIR) model \(^23,27\),

\[
dS = \lambda - (\beta t + \rho t) S - \delta S, \\
dE = (\beta i) S - (\nu t + \delta) E, \\
dI = \nu E - (\gamma + \delta) I + \rho I, \\
dR = \delta R,
\]

where \(N\) is the total number of population. The variables \(S, E, I, R\) denote the instantaneous size of the population related to the model. The total number of population susceptible to the disease at any instant of time is denoted by \(S\) while the number of population which is being exposed is denoted by \(E\). The variables \(I\) denotes the size of population at a given instant, which has been infected (fallen sick). The variable \(R\) denotes the combined size of population which
Fig. 1 The compartmental diagram for the model.

Table 2 Base-level average rates and total population.

| Symbol | Description                      | Value                       |
|--------|----------------------------------|-----------------------------|
| $\beta_t$ | transmission rate               | variable                    |
| $\rho_t$ | asymptomatic transmission rate  | variable                    |
| $\nu_t^{-1}$ | incubation period              | variable ($\leq 14$ days)  |
| $\gamma^{-1}$ | mean infection period         | $\sim 21$ days             |
| $\delta$ | death rate                      | $0.062$ day$^{-1}$          |
| $N$     | total population                | 1371.36 million             |

has either recovered from the disease or has succumbed to the disease i.e. the removed population. Naturally

$$S(t) + E(t) + I(t) + R(t) = N = \text{const.}$$  \hspace{1cm} (5)

In the above equations, $\beta_t$ is the rate of exposure of the population to the virus, $\rho_t$ is the rate of infection through the asymptomatic population, $\nu_t$ is the rate at which the exposed population contracts the disease i.e. the incubation period, $\gamma$ is the rate at which the population is removed from the infected population, and $\delta$ is the average death rate of the population (not due to the disease). The parameter $\lambda$ represents the import to the susceptible population from outside. The subscript ‘$t$’ in the rates indicate that the corresponding rates depend on time. All rates are expressed in terms of persons per day. The removed group includes removal of the population from the infected group either by recovery or by death. Note that at the limit $\nu_t \to \infty, \rho_t \to 0, \beta_t = \text{const.}$, the model reduces to the classic SIR model. The schematic compartmental diagram for the model is shown in Fig.1.

The disease free equilibrium (DFE) state of the system is given by $I(t) = E(t) = R(t) = 0$ and $\lambda = \delta S$, which is used to find the basic reproduction number $R_0$.

2.3 Modeling the rates

The base-level average values of the rates for India are tabulated in Table 2. As we have argued that at the start of the pandemic, the incubation period is almost zero as most number of population start at the infected compartment. In time, however, the incubation period will be applied to all the exposed population, who are being monitored. In the same way, at the beginning, the infection through asymptomatic population would be very small unless there is a large scale of migration of people. This always happens, but not at the beginning. Again, if we consider the effect of the nationwide lockdown, which most countries have resorted to, we can safely assume that the disease transmission rate constantly decreases, eventually becoming zero as people become accustomed to social distancing norms.

Based on these observation, we propose the following ansatzes,

$$\beta_t = \alpha t e^{-\frac{t}{\tau}},$$  \hspace{1cm} (6)

$$\rho_t = \eta \text{erf}(\rho_0 t),$$  \hspace{1cm} (7)

$$\nu_t = \nu_0 e^{-\frac{t}{\tau}} + \nu_f.$$  \hspace{1cm} (8)
In the first expression, $\alpha_t$ is a piecewise function, starting at the basic equivalent value for the classic SIR model at $t = 0$ and $\zeta_t$ is piecewise function which basically represents at what rate the transmission rate is decreasing. The piecewise property is necessary as at different times, we see different reasons which change the behavior of the pandemic. A sudden influx of migrants or a sudden uncontrolled congregation can be modeled with the piecewise behavior as and when needed. In the second expression, we assume that the asymptomatic transmission rate starts from a very low value $\rho_0 \sim 0$, gradually approaching a constant rate $\eta_1$, which is again a piecewise function and can be adjusted to suit sudden unexpected behavior of the pandemic. In the third expression, we assume that the incubation period approaches a final value of $\nu_f \approx 1/14$ days$^{-1}$ as the pandemic progresses. The quantity $\nu_0$ represents the inverse of initial incubation period at which the pandemic starts, which is a very large number, quickly decreasing determined by the rate $\varphi$.

3 The equilibria

The equilibria of the system, Eqs.\(^{(14)}\) is given by

$$\begin{pmatrix} \frac{\lambda}{\delta} \end{pmatrix}, \begin{pmatrix} 0 \end{pmatrix}, \begin{pmatrix} 0 \end{pmatrix}, 0, 0 \right)$$

where $R_0$ is the basic reproduction number (see the next section), which is given by

$$R_0 = \frac{\beta_1 \nu + \rho (\nu + \delta)}{(\gamma + \delta)(\delta + \nu)}$$

which we derive in the next section. Besides, we have also used the constraint given by \(^{(5)}\). The first equilibrium point is the disease free equilibrium (DFE) and the other one is the endemic equilibrium (EE) and it is amply clear that the system can reach EE only when $R_0 > 1$.

3.1 The basic reproduction number $R_0$

As we know that the basic reproduction number $R_0$ is defined as the average number of secondary cases that would be generated by a primary case in a totally susceptible population \(^{(28)}\). It provides an overall measure of transmission of an infectious disease within a given population, which is dependent on the transmission coefficient of the disease as well as on the average duration of infectiousness \(^{(28)}\). For $R_0 > 1$, the disease free equilibrium (DFE) of an infectious disease is unstable while for $R_0 < 1$, the DFE is asymptotically stable \(^{(76)}\).

In order to compute $R_0$, we consider our two infected compartments $E$ and $I$ of the model, we re-structure our model in terms of the rate of appearance of new infections $F_i(x)$ and the rate of transfer of individuals $V_i(x)$ \(^{(76)}\),

$$\dot{x}_i = F_i(x) - V_i(x), \quad i = 1, \ldots, n$$

with $n = 4$ and $x = (E, I, S, R)^t$. We note that the rate of transfer can further be divided into two transfer rates $V_i = V_i^- - V_i^+$, with $V_i^\pm$ denoting transfer into the $i$th compartment and out of it. The components are thus

$$F = \begin{pmatrix} \beta E S N, \rho I S N, 0, 0 \end{pmatrix}^t,$$

$$V = \begin{pmatrix} (\nu + \delta)E, -\nu I E + (\gamma + \delta)I, -\lambda + (\beta E + \rho I) IS N + \delta S, -\gamma J + \delta R \end{pmatrix}^t.$$  

The basic reproduction number $R_0$ is thus the leading eigenvalues of the matrix $FV^{-1}$, where

$$F = \left( \frac{\partial F_i}{\partial x_j} \right)_{x_{DFE}}^t, \quad V = \left( \frac{\partial V_i}{\partial x_j} \right)_{x_{DFE}}, \quad i, j = 1, 2, 4.$$
3.2 Endemic equilibrium

As shown at the beginning of the section, the EE of the system is given by the second solution of Eq. (9), which is expressed in terms of the basic reproduction number $R_0$, which can be reached only when $R_0 > 1$. It is interesting to see this EE limit of the variables which are the values when the epidemic is completely out of control, which is however, evaluated at the DFE state. For our model, we have

$$ F = \begin{pmatrix} 0 & \beta t \\ 0 & \rho t \end{pmatrix}, \quad V = \begin{pmatrix} \nu t + \delta & 0 \\ -\nu t & \gamma + \delta \end{pmatrix}, $$

where we have used the DFE state as $S/N = 1$. The basic reproduction number can now be found as

$$ R_0 = \frac{\beta t \nu t + \rho t (\nu t + \delta)}{(\gamma + \delta)(\delta + \nu t)}. $$

(16)
In this section, we try to map the pandemic in the federal states of India with our model. The Indian context posses a unique model to study this pandemic in terms of heterogenous groups of populations. As all the federal states of India are based on their respective cultures and languages, the states make somewhat homogenous groups of people over which these compartmental models are applicable. So, it is more appropriate to apply a model to individual states to get the overall picture for the whole country.

We assume that the asymptomatic infection rate starts from zero at the beginning of the pandemic and saturates at $n_t$, which is expressed as a fraction of the current disease transmission rate $\beta_t$. In the same way, the incubation period starts at a very large value $\gg 1$ at the beginning and saturates to $1/14 \text{ days}^{-1}$ as the case of new infections peaks up.
Fig. 3 The endemic equilibrium regions \( (R_0 > 1) \) in the \((\rho_t, \beta_t)\) space. The left panel is when both \( \rho_t \) and \( \beta_t \) are independent of each other and the right panel is when \( \rho_t = \xi \beta_t \), where \( \xi \) is a proportionality factor.

4.1 Optimization

The mathematical formulation to model the pandemic is set up as a nonlinear global optimization problem, where we try to solve the model equations numerically so that the deviation of the solution from the actual data is minimized globally with respect to the parameters \( \alpha_t, \zeta_t, \rho_0, \phi \). The parameter \( \nu_0 \) is fixed at \( 10^4 \), although any large positive number will do. The quantity which is minimized is given by

\[
\mathcal{R} = \sum_{t=1}^{N} \sum_{j=1}^{n} \left| x_j^{t} - \tilde{x}_j^{t} \right|^2 \tag{20}
\]

which is the residue (or global error) in fitting the data. This residue is the absolute squared difference between the numerically found values of the variables \( x = (E, I, S, R)' \) and their corresponding actual values \( \tilde{x} \) (from the data). The residue is accumulated over \( N \) sampling points on the data train denoting different time stamps over \( n \) number of solutions of the dynamical equations, which in this case is 4. The constrained global minimization is carried out through the derivative free Nelder-Mead algorithm \((30)\), though any other method such as trust-region-reflective can also be used. The minimization returns the best-fit parameters for the model.

4.2 Modeling at state level

We have tabulated the key parameters for each of these states in Table 4. In the table, the population \( P \) is indicated in millions as per latest data \((12)\), the gross domestic product (GDP) \((5)\) for the year 2020-21 \((P_{\text{GDP}})\) is shown in equivalent billion USD, \( P_{\text{Econ}} \) is the percentage of the population which has an average annual income less than about 380 USD \((18, 11)\), \( P_{\text{Urban}} \) is the percentage of the population living in urban areas, \( P_{\text{Child}} \) is the percentage of the children population (up to the age of 14 year), and \( P_{\text{Youth}} \) is the percentage of the youth population (from the age of 15 year to 59 year) \((2, 3)\). In Fig.4 we have plotted the total cumulative number of infectives \( I \), daily new infectives \( I_{\text{new}} \), and the basic reproduction number \( R_0 \) for each of the eleven federal states of India. In each of these figures, every row represents one state containing three panels of which the first panel shows the total infectives \( I \) as calculated from our numerical compartmental model (shown as dark black solid curve), which is superimposed with the actual data (shown as open red circles). The second panel shows the daily new infections as predicted by our model (shown as dark black curve), superimposed over the actual data (shown as yellow solid line). The third panel shows the corresponding \( R_0 \) for the particular state where the dotted line indicates the level \( R_0 = 1 \). The name of the state is abbreviated as per Table 4 and is indicated in each row in the first panel. The \( x \)-axis of each panel is the number of days elapsed since 14 March,
Fig. 4 (a) Total infectives $I$, daily new infectives $I_{\text{new}}$, and $R_0$. Every row represents one state with three panels of which the first panel shows the total infectives $I$ as calculated from our numerical compartmental model (shown as dark black solid curve), which is superimposed with the actual data (shown as open red circles). The second panel shows the daily new infections as predicted by our model (shown as dark black curve), superimposed over the actual data (shown as yellow solid line). The third panel shows the corresponding $R_0$, where the dotted line indicates the level $R_0 = 1$. (continued to the next page).
Fig. 4 (b) Total infectives $I$, daily new infectives $I_{\text{new}}$, and $R_0$ [See the explanatory texts in the caption of (a)] (continued to the next page).
2020. Note that this axis is different for each of these eleven states. In all these modeling, we have assumed that the transmission rate $\rho_t$ for the asymptomatic population saturates at 10% of the transmission rate i.e. $\rho_t = 0.1 \beta_t$.

We note that the first COVID-19 case was reported in India on 30 January, 2020. However, after that the occurrence of cases have either been sporadic or confined in certain localised households only. The major continuous detection of cases started in mid March, 2020 and in all our analysis we have taken the zeroth day as 14 March, 2020.

4.3 Transmission analysis — criticality

In case of an infectious disease like COVID-19, as the disease progresses, the number of people getting infected increases rapidly, a measure of which can be defined through the basic reproduction number $R_0$. In this section, we aim to find the day at which there is a sudden shift in the pattern of increase in the number of reported cases in various states of India. This sudden shift is present in almost every states, though we have here analysed only in 11 states which mostly contributed to the pandemic. We call this day as the critical day or the day of community transmission with an assumption that at around this day first major fluctuation takes place. We can also define this day as the day at which there is a shift from small case fluctuation of reported case to large scale fluctuations.

The Fig. 5 shows the transmission analysis of the disease in most 11 worst affected states. The left hand panels show the daily reported new cases $I_{\text{new}}$ vs. the total cumulative cases i.e. total infectives $I$, plotted as per real data (shown as the blue curve) and $5-30$ day mean $I_{av}$ (shown as the red curve). The right hand panels show the absolute deviation of the real data from its mean $\Delta I_{\text{new}}$, $\Delta I_{\text{new}} = |I_{\text{new}} - I_{av}|$, (21) which show a uniform behavioral pattern of the transmission. It is interesting to note that the fluctuation pattern of reported cases from its mean suddenly changes its behavior from a uniform small-scale fluctuation to a large-scale fluctuation (denoted by a vertical dashed line). The number $d_C$ denotes the position of this transition of behavior in number of days elapsed since the beginning of the data set i.e. 14 March, 2020. We have denoted this day as a critical day.

Table 4 Socio-economic parameters of 11 federal states of India.

| State       | Abbrev | $P^*$ (million) | $P^{GDP}$ (billion USD) | $P_{\text{Econ}}$ (%) | $P_{\text{Urban}}$ (%) | $P_{\text{Child}}$ (%) | $P_{\text{Youth}}$ (%) |
|-------------|--------|-----------------|-------------------------|------------------------|------------------------|------------------------|------------------------|
| Assam       | AS     | 35.61           | 57.0                    | 31.98                  | 14.08                  | 31.1                   | 62.80                  |
| Bihar       | BR     | 124.80          | 96.0                    | 33.74                  | 11.30                  | 37.30                  | 55.8                   |
| Delhi       | DL     | 18.71           | 120.0                   | 9.91                   | 97.50                  | 27.9                   | 66.2                   |
| Gujarat     | GJ     | 63.87           | 263.0                   | 16.63                  | 42.58                  | 27.8                   | 63.9                   |
| Karnataka   | KA     | 67.56           | 252.0                   | 20.91                  | 38.57                  | 25.7                   | 65.9                   |
| Maharashtra | MH     | 123.14          | 450.0                   | 17.35                  | 45.23                  | 27.2                   | 63.6                   |
| Odissa      | OD     | 46.36           | 75.0                    | 32.59                  | 16.68                  | 27.7                   | 63.0                   |
| Rajasthan   | RJ     | 81.03           | 158.0                   | 14.71                  | 24.89                  | 32.5                   | 60.2                   |
| Tamil Nadu  | TN     | 77.84           | 292.0                   | 11.28                  | 48.45                  | 23.4                   | 66.4                   |
| Uttar Pradesh | UP   | 237.88         | 250.0                   | 29.43                  | 22.78                  | 33.7                   | 59.5                   |
| West Bengal | WB     | 99.61           | 202.0                   | 19.98                  | 31.89                  | 25.3                   | 66.5                   |
Fig. 5 (a) Transmission analysis of 11 worst affected states. The left hand panels show the daily reported new cases $I_{\text{new}}$ vs. the total cumulative cases i.e. total infectives $I$, plotted as per real data (shown as the blue curve) and $5-30$ day mean $I_{\text{av}}$ (shown as the red curve). The right hand panels show the absolute deviation of the real data from its mean $\Delta I_{\text{new}}$ [see Eq. (21)], which show a uniform behavioral pattern of the transmission. The fluctuation pattern of reported cases from its mean, suddenly changes its behavior from a uniform small-scale fluctuation to a large-scale fluctuation (denoted by a vertical dashed line). The number $d_C$ denotes the position of this transition of behavior in number of days elapsed since the beginning of the data set i.e. 14 March, 2020. (continued to the next page).
Fig. 5 (b) Transmission analysis of 11 states[See the explanatory texts in the caption of (a)] (continued to the next page).
day $d_C$ which denotes the first major change of transmission rate in the respective state. We may very well call it as the day when the community transmission (or large scale transmission) starts. It is worth noting that following continuous reporting of positive COVID-19 cases, which started on 14 March, 2020, national level lockdown in the country was imposed in four phases (14)

1. 1st phase – from 25 March to 14 April, 2020 (up to 31st day)
2. 2nd phase – from 15 April to 3 May, 2020 (up to 50th day)
3. 3rd phase – from 4 May to 17 May, 2020 (up to 64th day)
4. 4th phase – from 17 May to 31 May, 2020 (up to 78th day)

As we see that on average, the criticality started in and around the end of the 4th level of lockdown i.e. near about 80th day (the average being 78.8 days). While five states, the criticality started around 100th day, in six other states it

Fig. 5 (c) Transmission analysis of 11 states. [See the explanatory texts in the caption of (a)]
started within 44th and 79th day. From these observations, we can see that these plots correctly captures the transmission characteristics of the pandemic in these major states.

4.4 Scatter plots — the socio-economic pattern

As mentioned before, in our present analysis, we have taken 11 most affected states of India. From population point of view, most of the Indian states have comparable populations with other individual countries. Within India, most of the states differ in terms of culture, economy, and demography. In this section, we aim to draw some relations in respect of spreading of the SARS-CoV-2 disease and different socio-economic parameters.

We show the scatter plots of the disease transmission parameters — $\beta_{C,N}$ and $d_{C}$ versus the socio-economic parameters as described in Table 4. $\beta_{C,N}$ are the disease transmission rates at the criticality i.e. on the critical day $d_{C}$ and at the end of the data interval ($N$th day), which is 31st August, 2020, respectively. In all the following analyses, we have taken three limiting values for the asymptomatic transmission rate $\rho_{t}$ — 10%, 20%, and 30% of the transmission rate $\beta_{t}$.

For fitting the data, we have used nonlinear Levenberg-Marquardt method with all data points weighted equally i.e. with a weight one. The goodness of fitting is denoted by the so called quality factor $Q$, which can be defined in terms of normalised incomplete gamma function $\Gamma(N_{DOF}/2, x^{2}/2)$:

$$Q = \frac{1}{\Gamma(N_{DOF}/2)} \int_{x^{2}/2}^{\infty} x^{N_{DOF}/2-1} e^{-x} dx \leq 1,$$

where $N_{DOF}$ is the “number of degrees of freedom”, which is defined as $N_{DOF} = N - M$, where $N$ is the number of independent data points and $M$ is the independent parameters of fitting. The chi-square, defined as

$$\chi^{2} = \sum_{i=1}^{N} \left[ \frac{y_{i} - f(x_{i})}{\sigma_{i}} \right]^{2},$$

with the nonlinear fitting model $y = f(x)$. The value of $Q$ is indicated in each of the scatter plots.
Fig. 7 $\beta_C$ vs. $P_{\text{Urban}}$ (Delhi omitted): The urban population also seems to drive the first major transmission. The more the urban population, the more quick the day of first major transmission $d_C$ is reached.

Fig. 8 $\beta_C$ vs. $P_{\text{GDP}}$: The more affluent is the population, the more is the transmission rate at the first major transmission. So, it is the purchasing power of the population which has driven the transmission.

Fig. 9 $\beta_C, \beta_N$ vs. $P_{\text{Econ}}$: $P_{\text{Econ}}$ is the % of population below a certain income denoted as Economy group. The lower is the value of $P_{\text{Econ}}$, the richer is the population.
4.4.1 Disease transmission and population age

In Fig.6, we show the critical transmission rate $\beta_C$ versus the youth and child populations in the selected states. Each row of the figure contains three columns of scatter plots which are respectively for 10, 20, and 30% asymptomatic transmission rate (see the beginning of this section). The scatter plots show a power law dependence of $\beta_C$ on the youth and child population. While $\beta_C$ seems to increase with the youth population, it decreases with the child population. This is quite understandable as the child population is effectively confined during the lockdown period and thereafter owing to closure of educational institutions, the youth population – primarily the working population begins to go out thus contributing to the growth of the disease. So, we can conclude that during the extended lockdown phase, the youth population seems to drive the pandemic.

4.4.2 Disease transmission and urban population

In Fig.7, we show how does the critical transmission rate $\beta_C$ depend on urban population. The scatter plots clearly show that the urban population is one of the major parameters which drives the pandemic. The asymptomatic transmission rate is modeled at 10, 20, and 30% of the primary transmission rate as before.

4.4.3 The critical day, disease transmission and economy

In Fig.8, we have shown the dependence of $\beta_C$ on the GDP of the respective state. The scatter plots suggest that the more affluent is the population (having more purchasing power), the more is the transmission rate at the first major transmission. So, it is the purchasing power of the population which has also driven the transmission.

Fig.9 contains two rows of which the first row shows the dependence of $\beta_C$ on the economy population $P_{\text{Econ}}$ and the second row shows the dependence of the end transmission rate $\beta_N$ (as on 31 August, 2020) $P_{\text{Econ}}$. So, the less is the $P_{\text{Econ}}$ for a state, the more is the people in state with more purchasing power and the richer is the population. This figure presents an interesting study. We see that during the national level lockdown, which lasted for about 68 days, the transmission of the disease was mainly driven by the affluent people which points to the fact that the population with less purchasing power stayed confined. The lower is the purchasing power of the population (i.e. more poorer people), lower is the transmission rate at the first major transmission and the longer it requires to have first major transmission (see Fig.10). It is interesting to note that this behavior is in contrast with the behavior at the later phase of the pandemic. As the pandemic progresses, the transmission rate becomes proportional to the $P_{\text{Econ}}$, indicating the pandemic fatigue of the low income group.

As shown in Fig.10, the richer is the population and the more is the urban population, the quicker it takes to reach the criticality (lower $d_C$), which corroborates our earlier observation about the affluent and urban people driving the early phase of the pandemic.
Fig. 11  \( I \) and \( I_{\text{New}} \) for India as analysed by the model based on the experiment period of 170 days (from 14 March to 31 August, 2020) and the projection till 400 days (18 April, 2021). The red coloured dotted parts in the panels shows the projection. The number inscribed in each row is the value of \( r \).
Table 5  Mean squared error $\epsilon_{\text{MSE}}$ (expressed per million) for the projected data for various values of $r$, $\rho_t = r\beta_t$.

| $r$   | $\epsilon_{\text{MSE}}$ |
|-------|------------------------|
| 5%    | $3.476 \times 10^{-4}$ |
| 7.5%  | $1.406 \times 10^{-4}$ |
| 10%   | $3.033 \times 10^{-4}$ |
| 12.5% | $2.553 \times 10^{-4}$ |
| 15%   | $2.231 \times 10^{-4}$ |

5 The overall scenario for the whole country

We shall now apply our model for the entire country. In what follows, we present the result with the data available from 14 March to 31 August, 2020, for a duration of 170 days in Fig.11. Each row in Fig.11 has three panels. The first two panels show the results for the cumulative infectives $I$ and daily new infectives $I_{\text{New}}$ for the period of 170 days. The model uses these 170-day data to model the pandemic. The real data for this period is shown by the thick orange solid line whereas the data beyond 170 days up to 224 days (for a period of 54 days) is represented by a dotted red line. The numerically calculated curve is shown by the black solid curve. So, the dotted portion of the data shows the projection of the pandemic by the model based on its analysis during the past 170 days. The third panel in each row shows the reproduction number $R_0$. In each row, we have indicated the actual date for the entire period and the thin vertical line marks the 170th day i.e. 31 August, 2020 till when the data were used for the model to predict the rest of the figure.

There are five rows in Fig.11, each row is for the limiting asymptomatic rate $\rho_t = r\beta_t$, $r = 5, 7.5, 10, 12.5,$ and 15%. The mean square error $\epsilon_{\text{MSE}}$, defined as

$$\epsilon_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2,$$

expressed per million of population for each value of $r$ tabulated in Table.5. In the above expression, $Y_i$ is the actual data and $\hat{Y}_i$ is the corresponding estimated value predicted by the model. As one can see that the $\epsilon_{\text{MSE}}$ is minimum for $r = 7.5\%$. So, we can safely conclude that the optimum asymptomatic transmission rate of the disease for India is $< 10\%$.

6 Summary and Conclusions

To summarise, taking into consideration of asymptomatic carriers for COVID-19 pandemic, an SEIR model for India is proposed with time varying transmission rates and incubation period, as it is believed that the asymptomatic population seems to play an important role in the propagation of the SARS-CoV-2 disease. With the assumption that the asymptomatic transmission rate ($\rho_t$) can be written as a fraction of the symptomatic transmission rate ($\beta_t$), the disadvantage of the non-availability of asymptomatic infective data can be incorporated in our model without taking an extra compartment for asymptomatic population. We have also incorporated time varying incubation period for the virus in our model.

Numerical results obtained from the model are first fitted for 11 Indian states, which are the major contributors to the pandemic and also fitted for the entire country. The mathematical modelling is based on data for India up to 31 August 2020, which has been used to model the behavior of the pandemic for a total of 400 days (14 March 2020 – 18 April 2021). We observed that our predicted results agree well with the data up to 54 days, beyond the end of the data used in the modeling. The mean squared error between the predicted data and the actual data is minimum for the asymptomatic transmission rate of 7.5%.

We also proposed a way to pinpoint the time when first major transmission occurs (we called it critical day) for a particular state and transmission rate at that day is calculated and compared with various socio-economic parameters
for the respective states. Variation of the transmission rates at the critical day with youth and children population was studied and it is found that, state with higher youth population has larger transmission rate on the critical day. If a state has higher child population, the critical transmission rate has been found to be lower for that state. We also find that the higher the urban population of a state, the higher is the critical transmission rate and the earlier is the day on which a major transmission occurs. States with higher GDP (more purchasing power) experiences higher rate of transmission at the critical day. These are some interesting results of this work that reflect the correlation between the behavior of the pandemic with socio-economic and demographic structure of Indian states.

We also note that as per the results of our analysis, at the start of the national level lockdown, the pandemic was driven primarily by the people with higher income while as the pandemic is progressing, it has been gradually found to be driven by the people with relatively lower income, demonstrating the so called pandemic fatigue. For India, we have found an asymptomatic transmission rate of $\sim 7.5\%$ to be an optimum rate for the best fit parameters. Since our model has successfully predicted the progress of the pandemic up to 54 days, we believe that the projection of the model up to 18 April 2021 can be used as a guidance to follow further preventive and protective measures until next summer. As is seen in Fig.11, the value of $R_0$ falls to $\sim 0.5$ by the end of April 2021, when the disease can be taken as contained as $R_0$ falls sufficiently below unity. Our equilibria analysis shows that in India, an endemic equilibrium can reach when the the infectives can reach about 415 million, which is very far from the cumulative total of about 10.5 million cases (as of mid-January 2021). Thus, we can safely assume that India will never reach an EE state, now that the vaccination program has already started.

To conclude, we have shown how the COVID-19 pandemic can be related to the socio-economic status and demography of a particular region, of India in particular. The results of this model can be used for early dynamic preventive measures for similar situations in the developing countries.

Funding

This work is carried out with a support from the SERB-DST (India) research grant no. CRG/2018/002971. Necessary computational facilities, in part, are provided through institutional FIST support (DST, India).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Census of India 2011 (2011). URL https://censusindia.gov.in/2011Census/C%2014_25052013_NRM.pdf
2. Population composition, chapter 2 (2011). URL https://censusindia.gov.in/vital_statistics/SRS_Report/9Chapt%202%20-%2020%202011.pdf
3. Provisional Population Totals, Census of India 2011, vol. 1 (2011). URL https://censusindia.gov.in/2011-prov-results/paper2/data_files/india/paper2_1.pdf
4. Sample Registration System, Baseline Survey 2014 (2014). URL https://censusindia.gov.in/vital_statistics/BASELINE%20TABLE%2007062016.pdf
5. Analysis of state budgets (2020). URL https://www.prsindia.org/parliamenttrack/budgets/state(2020)
6. Countries in the world by population (2020). URL https://www.worldometers.info/world-population/population-by-country
7. COVID-19 India Org Data Operations Group (2020). URL https://api.covid19india.org
8. Johns Hopkins University, Coronavirus Resource Center (2020). URL https://coronavirus.jhu.edu
9. Ministry of Health and Family Welfare (2020). URL https://www.mohfw.gov.in
10. National Accounts Data (2020). URL http://mospi.nic.in/download-tables-data
11. SDG India index, NITI Aayog (2020). URL https://sdgindiaindex.niti.gov.in/#/ranking
12. State/UT wise Aadhaar Saturation (2020). URL https://uidai.gov.in/images/state-wise-aadhaar-saturation.pdf
13. Anastassopoulou, C., Russo, L., Tsakris, A., Siettos, C.: Data-based analysis, modelling and forecasting of the covid-19 outbreak. PLoS One 15(3), e0230405 (2020)
14. Bagal, D.K., Rath, A., Barua, A., Patnaik, D.: Estimating the parameters of susceptible-infected-recovered models of covid-19 cases in India during lockdown periods. Chaos Solitons Fractals 140, 110154 (2020)
15. Barman, M.P., Rahman, T., Bora, K., Borghain, C.: Covid-19 pandemic and its recovery time of patients in india: a pilot study. Diabetes Metab Syndr 14, 1205–1211 (2020)
16. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci 180(1-2), 29–48 (2002)
17. Fanelli, D., Piazza, F.: Analysis and forecast of covid-19 spreading in china, italy and france. Chaos Solitons Fractals 134, 109761 (2020)
18. Fernandez, B.: Poor practices: contestations around ‘below poverty line’ status in india. Third World Q. 31(3), 415–430 (2010)
19. Giordano, G., et al.: Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. Nat. Med. 26, 855–860 (2020)
20. He, S., Peng, Y., Sun, K.: Seir modeling of the covid-19 and its dynamics. Nonlinear Dyn 101, 1667–1680 (2020)
21. Kakar, A., Nundy, S.: Covid-19 in india. J. R. Soc. Med. 113(6), 232–233 (2020)
22. Kaxiras, E., Neofotistos, G., Angelaki, E.: The first 100 days: Modeling the evolution of the covid-19 pandemic. Chaos Solitons Fractals 138, 110114 (2020)
23. Li, M.Y.: An introduction to mathematical modeling of infectious diseases. Springer International Publishing AG (2018)
24. Lin, Q., et al.: A conceptual model for the coronavirus disease 2019 (covid-19) outbreak in wuhan, china with individual reaction and governmental action. Int. J. Infect. Dis. 93, 211–216 (2020)
25. Linton, N.M., et al.: Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. J. Clin. Med. 9, 538 (2020)
26. Mahajan, A., Sivadas, N.A., Solanki, R.: An epidemic model sipherd and its application for prediction of the spread of covid-19 infection in india. Chaos Solitons Fractals 140, 110156 (2020)
27. Martcheva, M.: An introduction to mathematical epidemiology. Springer US (2015)
28. Miller, E.: The Vaccine Book, chap. 1, pp. 37–50. Academic Press (2003)
29. Nabi, K.N.: Forecasting covid-19 pandemic: A data-driven analysis. Chaos Solitons Fractals 139, 110046 (2020)
30. Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. 7(4), 308–313 (1965)
31. Pal, S.C., et al.: Threats of unplanned movement of migrant workers for sudden spurt of covid-19 pandemic in india. Cities (2020)
32. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes. Cambridge university press (2007)
33. Quadri, S.A.: Covid-19 and religious congregations: implications for spread of novel pathogens. Int. J. Infect. Dis. 96, 219–221 (2020)
34. Sahoo, B.K., Sapra, B.K.: A data driven epidemic model to analyse the lockdown effect and predict the course of covid-19 progress in india. Chaos Solitons Fractals 139, 110034 (2020)
35. Samui, P., Mondal, J., Khajanchi, S.: A mathematical model for covid-19 transmission dynamics with a case study in india. Chaos Solitons Fractals 140, 110173 (2020)
36. Sarkar, K., Khajanchi, S., Nieto, J.J.: Modeling and forecasting the covid-19 pandemic in india. Chaos Solitons Fractals 139, 110049 (2020)
37. Sengupta, S., Jha, M.K.: Social policy, covid-19 and impoverished migrants: challenges and prospects in locked down india. The International Journal of Community and Social Development 2(2), 152–172 (2020)
38. Young, P.: Everything You Wanted to Know About Data Analysis and Fitting but Were Afraid to Ask, 1 edn. Springer International Publishing (2015)
39. Zhong, L., Mu, L., Li, J., Wang, J., Yin, Z., Liu, D.: Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model. IEEE Access 8, 51761–51769 (2020)