COVID-19 outbreak in India: an SEIR model-based analysis

Debashis Saikia · Kalpana Bora · Madhurjya P. Bora

Abstract We present a modelling and analysis of the COVID-19 outbreak in India with an emphasis on the socio-economic composition, based on the progress of the pandemic (during its first phase—from March to August 2020) in 11 federal states where the outbreak is the largest in terms of the total number of infectives. Our model is based on the susceptible-exposed-infectives-removed (SEIR) model, including an asymptomatic transmission rate, 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 42 days into the months of September and October 2020, based on the past data. Overall, we have presented a projection up to 351 days (till February 2021) for India and we have found that our model is able to predict correctly the first phase of the pandemic in India with correct projections of the peak of the pandemic as well as daily new infections. 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 behaviour of the pandemic with demographic and socio-economic parameters. Towards the end, we have modelled the available data with the help of logistic equations and compare this with our model. The results of this work can be used as a future guide to follow in case of similar pandemics in developing countries.

Keywords COVID-19 · SEIR model · Epidemiology

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 in 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 [34,36]. As COVID-19 is a pandemic, which is still evolving, 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 [2,9,11,12,16,18,21,25,34,37,44]. As of mid-March 2021, total global cases of the pandemic have reached more than 120 million [14].

In India, the first COVID-19 case was confirmed on the 30 January 2020 [8]. 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 [15,32] along with the city of Mumbai. One unique cause which fuelled the spreading of the disease in certain states such as Bihar, Assam,
Table 1 Comparison of population of some of the Indian states [7]

| 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 |

West Bengal, and Odisha was the movement of migrant workers during May 2020 [28,39].

The Indian scenario as far as the pandemic is concerned is quite intriguing. Some of the major facts which justify an India-specific modelling are the huge population size and its density, the status of the economy, extreme diversity in the 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 behaviour of the population can vary widely from state to state as the lifestyle and culture, which govern the social behaviour, are very different as one travels across the length and breadth of the country. The same can be said for the economy. The gross domestic product (GDP) stands at a staggering 450 billion USD for the state of Maharashtra to a minimal 1.2 billion USD for the union territory of Andaman and Nicobar Islands [1,26]. 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 [6]. It is also important to note that the lifestyle 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 [35], about 30% of the Indian population are vegetarian. As this socio-economic behaviour is closely related to the progress of the pandemic, the response of the state population to the pandemic is worth studying. The population of many Indian states is even comparable to some countries. Table 1 shows how the populations of certain Indian states can be compared with some countries of the world [7].

In this work, we have constructed a compartmental non-autonomous disease model based on the well-known susceptible-exposed-infectives-removed (SEIR) model [17,22]. 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. At this point, we must caution the reader that this is a pandemic that is still evolving and even during the course of writing this paper, the pandemic has evolved beyond the time that was originally intended as a prediction limit for this work. While, at the time of initiating this work, the surge of the infections in India was yet to attain its peak, at present, we are witnessing the second wave of the pandemic, which started from about March 2021. However, we have found that the model described here is able to predict the first phase of the pandemic, almost in entirety, well beyond its intended prediction limit—about 180 days beyond the analysed 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 [15,32]. 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 behaviour the way the pandemic has advanced. We have limited our analy-
sis of the data to the end of August 2020—to the 170th day into the pandemic from the day of first detected infection, and all the predictions beyond the 170th day are based on the analysis of data only up to this date. We have limited the analysis of data up to this date, 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 in Sect. 2. Next, we present an equilibria analysis for our model and calculated the basic reproduction number $R_0$ in Sect. 3. Then, we analyse the progress of the pandemic for eleven Indian states in Sect. 4. In Sect. 5, we provide an overall scenario for the whole country and provide our projections of the pandemic. In Sect. 6, we compare the results obtained from our model with the logistic models. Finally, we summarise and conclude with the novel ideas used and the main results obtained in this work in Sect. 7. We would also like to point to the fact that all the figures up to Sect. 5 are based on the data that were available till about mid-October 2020 (212th day into the pandemic). In Sect. 6, however, we have used the latest data, available till March 2021.

2 The model

2.1 The limitations on data

It is worth noting that though the gross official COVID-19 data are maintained by the Ministry of Health and Family Welfare (MoHFW) in India [24], 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 groups, 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 modelled the pandemic on the basis of data what is available in the public domain. For modelling the state data, we have use crowdsourced data, which is maintained by https://www.covid19india.org [8]. It should be noted that this dataset closely matches the official data provided by the MoHFW and the data maintained by John Hopkins’ COVID database [14].

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 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 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 deaths 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 an asymptomatic carrier.

As almost all nations have resorted to a lockdown or other restrictive measures once they have detected that there is a section of the 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 the 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 [17,22],

\[
\frac{dS}{dt} = \lambda - (\beta_t + \rho_t) IS \frac{I}{N} - \delta S, \\
\frac{dE}{dt} = \beta_t IS \frac{I}{N} - (\nu_t + \delta) E, \\
\frac{dI}{dt} = \nu_t E - (\gamma + \delta) I + \rho_t IS \frac{I}{N}, \\
\frac{dR}{dt} = \gamma I - \delta R,
\]

where \( N \) is the total population. The variables \( S, E, I, R \) denote the instantaneous size of the population related to the model. The total number of the 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 variable \( I \) denotes the size of the population at a given instant, which has been infected (fallen sick). The variable \( R \) denotes the combined size of the population which has either recovered from the disease or has succumbed to the disease, i.e. the removed population. Naturally

\[ S(t) + E(t) + I(t) + R(t) = N = \text{const.} \tag{5} \]

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 \( \cdot t \) in the rates indicates that the corresponding rates depend on time. All rates are expressed in terms of persons per day. The removed group includes the population, which gets removed from the infected group either by recovery or by death. Note that at the limit \( \nu_t \rightarrow \infty, \rho_t \rightarrow 0, \beta_t = \text{const.}, \) 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 Modelling 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 the 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, in the beginning, the infection through the 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 observations, we propose the following ansätze,

\[
\beta_t = \alpha_t e^{-\zeta t}, \\
\rho_t = \eta_t \text{erf}(\rho_0 t), \\
\nu_t = \nu_0 e^{-\psi t} + v_f. \tag{8}
\]

In the first expression, \( \alpha_t = \alpha_0 \hat{\alpha}(t) \) is a piecewise function, starting at the basic equivalent value for the classic SIR model at \( t = 0 \) and \( \zeta \) is \( \zeta_0 \hat{\zeta}(t) \) is a 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 behaviour of the pandemic. A sudden influx of migrants or a sudden uncontrolled congregation can be modelled with the piecewise behaviour 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 \( \eta_t \), which is again a piecewise function and can be adjusted to suit sudden unexpected behaviour of the pandemic. In the third expression, we assume that the incubation period approaches a final value of \( v_f \approx 1/14 \text{ days}^{-1} \) [19] as the pandemic progresses. The quantity \( \nu_0 \) represents the inverse of the initial incubation period at which the pandemic starts, which is a very large number, quickly decreasing determined by the rate \( \varphi \).
COVID-19 outbreak in India: an SEIR model-based analysis

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                        |
| $\nu^{-1}$ | Incubation period         | variable ($\leq 14$ days) [19] |
| $\gamma^{-1}$ | Mean infection period     | $\sim 21$ days [5]              |
| $\delta$ | Death rate                | $0.062$ day$^{-1}$ [36]         |
| $N$    | Total population          | $1371.36$ million               |

3 The equilibria

The equilibria of the system, Eqs. (1–4), are given by

\[
(S, E, I, R) = \begin{cases} 
\left( \frac{\lambda}{\delta}, 0, 0, 0 \right) \\
\left( \frac{\lambda}{R_0\delta - R_0(\delta + \nu_t)(\beta_t + \rho_t)}, \frac{\lambda(R_0 - 1)}{\beta_t + \rho_t}, \frac{\gamma\lambda(R_0 - 1)}{\delta(\beta_t + \rho_t)} \right) 
\end{cases},
\]

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

\[
R_0 = \frac{\beta_t\nu_t + \rho_t(\nu_t + \delta)}{(\gamma + \delta)(\delta + \nu_t)},
\]

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 [23]. 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 [23]. 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 [41].

In order to compute $R_0$, we consider our two infected compartments $E$ and $I$ of the model, and we restructure our model in terms of the rate of appearance of new infections $\mathcal{F}_i(x)$ and the rate of transfer of individuals $\mathcal{V}_i(x)$ [41],

\[
\dot{x}_i = \mathcal{F}_i(x) - \mathcal{V}_i(x), \quad i = 1, \ldots, n
\]

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

\[
\mathcal{F} = \left( \frac{\beta_t IS}{N}, \frac{\rho_t IS}{N}, 0, 0 \right)',
\]

\[
\mathcal{V} = \left( (\nu_t + \delta)E, -\nu_t E + (\gamma + \delta)I, -\lambda + (\beta_t + \rho_t)\frac{IS}{N} + \delta S, -\gamma I + \delta R \right)',
\]
The basic reproduction number $R_0$ is thus the leading eigenvalues of the matrix $F V^{-1}$, where

$$ F = \left( \frac{\partial f_i}{\partial x_j} \right)_{\text{DFE}}, \quad V = \left( \frac{\partial V_i}{\partial x_j} \right)_{\text{DFE}}, \quad i, j = 1, 2 $$

evaluated at the DFE state. For our model, we have

$$ F = \begin{pmatrix} 0 & \beta_1 \\ 0 & \rho_1 \end{pmatrix}, \quad V = \begin{pmatrix} \nu_1 + \delta & 0 \\ -\nu_1 & \gamma + \delta \end{pmatrix}, $$

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

$$ R_0 = \frac{\beta_1 \nu_1 + \rho_1 (\nu_1 + \delta)}{(\gamma + \delta)(\delta + \nu_1)}. $$

### 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, not expected to reach. The EE values of the variables in terms of the basic rates are given by,

$$ E_{EE} = \frac{\beta_1 \lambda}{(\beta_t + \rho_t)} \left[ \frac{1}{(\delta + \nu_t)} - \frac{(\gamma + \delta)}{\beta_t \nu_t + \rho_t (\delta + \nu_t)} \right], $$

$$ I_{EE} = \frac{\lambda}{(\beta_t + \rho_t)} \left[ \frac{\rho_t}{(\gamma + \delta)} + \frac{\beta_1 \nu_t}{(\delta + \nu_t)(\gamma + \delta)} - 1 \right], $$

$$ R_{EE} = \frac{\gamma}{\delta} I_{EE}. $$

From the above expressions, it can be shown that a sufficient condition for the system to reach an endemic equilibrium (EE) is $\rho_t > \gamma + \delta$ or when the asymptomatic transmission rate overtakes the combined value of removal and death rates.

As mentioned at the beginning, it is logical to assume that the asymptomatic transmission rate $\rho_t$ is proportional to the usual transmission rate $\beta_t$ of disease, though a completely independent asymptomatic transmission rate is not impossible to have. Various endemic equilibrium values of the variables in the two limiting cases—when $\beta_t \gg 1$ (independent of $\rho_t$) and $\rho_t \gg 1$, are shown in Table 3. In the limit of a very high asymptomatic transmission rate, we have no exposed population left in the system as people become infected (asymptomatically) without being considered as exposed. The incubation rate $\nu_t$, mean infection rate $\gamma$, and the total number of population (for India) are as per Table 2.

In Fig. 2, we have shown the values of the exposed population, infectives, and removed population approaching the EE state when $R_0 > 1$. The top panel shows the variable when the asymptomatic transmission rate $\rho_t$ stays at 10% of the transmission rate $\beta_t$, i.e. at all times $\rho_t = 0.1 \beta_t$. In the middle panel, we show the variables when $\rho_t$ is kept independent of $\beta_t$ and has a value of $10^{-4}$ day$^{-1}$. The bottom panel shows the variables when $\beta_t$ is kept independent of $\rho_t$ and has a value of $10^{-4}$ day$^{-1}$. The dotted lines in the figure are the limiting values of each variable as given in Table 3. In Fig. 3, we show the regions of EE in the ($\rho_t$, $\beta_t$) space, which is the darkened region with $R_0 > 1$. While the left panel in the figure is for independent $\rho_t$ and $\beta_t$, the right panel is when $\rho_t = \xi \beta_t$, where $\xi$ is a proportionality factor. As we can see from the figure that for a country like India, for the disease to reach an endemic state, the asymptomatic transmission rate $\rho_t \geq 0.54 \beta_t$ (the red coloured straight lines in the panels of Fig. 3) or the asymptomatic transmission rate should be as high as about 54% of the normal transmission rate.

### 4 Case examples

In this section, we try to map the pandemic in the federal states of India with our model. The Indian context possesses a unique model to study this pandemic in terms of heterogeneous groups of populations. As all the federal states of India are based on their respective cultures and languages, the states make somewhat homogeneous 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 $\eta_t$, 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 about 1/14 days$^{-1}$ as the case of new infections peaks up.
4.1 Optimisation

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

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

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 minimisation is carried out through the derivative-free Nelder–Mead algorithm [27], though any other method such as trust-region-reflective can also be used. The minimisation returns the best-fit parameters for the model.

It should, however, be noted that an unconstrained minimisation is not possible due to the complexity of the fitting functions which have piecewise time-dependent functions as there may be several local minima and the minimisation may converge to any one of

The dotted line in each panel represents the limiting value when the variable at abscissa becomes $\gg 1$.
Table 3 The endemic equilibrium (EE) values of the variables in two limits. The numbers are shown in millions

|                | \(\beta_t \gg 1\) | \(\rho_t \gg 1\) |
|----------------|-------------------|------------------|
| \(E_{\text{EE}}\) | \(\frac{\lambda}{\delta + \nu_t}\) | 637.23           | 0                |
| \(I_{\text{EE}}\) | \(\frac{\nu_t}{(\gamma + \delta)} E_{\text{EE}}\) | 415.22           | \(\frac{\lambda}{\gamma + \delta}\) |
| \(R_{\text{EE}}\) | \(\frac{\gamma}{\delta} I_{\text{EE}}\) | 318.91           | \(\frac{\gamma}{\delta} I_{\text{EE}}\) |

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.

Table 4 Convergence region for optimisation

| Constraints | Variables | MH          | GJ           |
|-------------|-----------|-------------|--------------|
| \(10^{-2} > \alpha_0 > 10^{-4}\) | \(\alpha_0\) | \(1.14 \times 10^{-3}\) | \(1.05 \times 10^{-3}\) |
| \(10^{-1} > \zeta_0 > 10^{-3}\) | \(\zeta_0\) | \(6.57 \times 10^{-3}\) | \(8.49 \times 10^{-3}\) |
| \(10^{-1} > \varphi > 10^{-3}\) | \(\varphi\) | \(5.47 \times 10^{-2}\) | \(6.18 \times 10^{-2}\) |
| \(10^{-1} > \rho_0 > 10^{-3}\) | \(\rho_0\) | \(7.26 \times 10^{-3}\) | \(1.08 \times 10^{-2}\) |
| \(10^{-1} > \xi > 10^{-3}\) | \(\xi\) | \(1.48 \times 10^{-2}\) | \(1.79 \times 10^{-2}\) |

4.2 Modelling at state level

We have tabulated the key parameters for each of these states in Table 5. In the table, the population \(P\) is indicated in millions as per the latest data [40], the gross domestic product (GDP) [1] 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 [10,38], \(P_{\text{Urban}}\) 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 years), and \(P_{\text{Youth}}\) is the percentage of the
The convergence of the constrained global minimisation for two states—Maharashtra (MH) and Gujarat (GJ) for the asymptomatic rate of 10%.

Table 5  Socio-economic parameters of 11 federal states of India

| State         | Abbrev | $P$ (million) | $P_{GDP}$ (billion USD) | $P_{Econ}$ (%) | $P_{Urban}$ (%) | $P_{Child}$ (%) | $P_{Youth}$ (%) |
|---------------|--------|---------------|--------------------------|----------------|----------------|----------------|----------------|
| Assam         | AS     | 35.61         | 57.0                     | 31.98          | 14.08          | 31.1           | 62.8           |
| Bihar         | BR     | 124.80        | 96.0                     | 33.74          | 11.30          | 37.3           | 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           |
| Odisha        | 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           | 66.1           |
| Uttar Pradesh | UP     | 237.88        | 250.0                    | 29.43          | 22.28          | 33.7           | 59.5           |
| West Bengal   | WB     | 99.61         | 202.0                    | 19.98          | 31.89          | 25.5           | 66.3           |

We note that the first COVID-19 case was reported in India on 30 January 2020. However, after that the occurrence of cases has 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 analyses, we have taken the zeroth day as 14 March 2020.
Fig. 5  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 $R_0 = 1$. b Total infectives $I$, daily new infectives $I_{\text{new}}$, and $R_0$ [See the explanatory texts in the caption of (a)] c Total infectives $I$, daily new infectives $I_{\text{new}}$, and $R_0$. [See the explanatory texts in the caption of (a)]
Fig. 5 continued
4.3 Transmission analysis: criticality

In the 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 11 states which mostly contributed to the pandemic. We call this day 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-scale fluctuations of reported cases to large-scale fluctuations.

Figure 6 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 \( I_{\text{av}} \) (shown as the red curve). The right-hand panels show the absolute deviation of the real data from its mean \( \Delta I_{\text{new}} \),

\[
\Delta I_{\text{new}} = |I_{\text{new}} - I_{\text{av}}|,
\]

which show a uniform behavioural pattern of the transmission. It is interesting to note that the fluctuation pattern of reported cases from its mean suddenly changes its behaviour 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 behaviour 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 \( d_C \) which denotes the first major change of transmission rate in the respective states. 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 [4]

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 in five states, the criticality started around the 100th day, in six other states it started within the 44th and the 79th day. From these observations, we can see that these plots correctly capture 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 a 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 5. \( \beta_{C,N} \) 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 31 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 the nonlinear Levenberg–Marquardt method [30] 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 \( \Gamma(N_{\text{DOF}}/2, \chi^2/2) \) [43]

\[
Q = \frac{1}{\Gamma(N_{\text{DOF}}/2)} \int_0^\infty x^{N_{\text{DOF}}/2-1} e^{-x} \, dx \leq 1,
\]

where \( N_{\text{DOF}} \) is the ‘number of degrees of freedom’, which is defined as \( N_{\text{DOF}} = N_p - M \), where \( N_p \) is the number of independent data points and \( M \) is the independent parameters of the fitting. The Chi-square is defined as

\[
\chi^2 = \sum_{i=1}^{N_p} \frac{(y_i - f(x_i))}{\sigma_i}^2.
\]
Fig. 6  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 behavioural pattern of the transmission. The fluctuation pattern of reported cases from its mean suddenly changes its behaviour from a uniform small-scale fluctuation to a large-scale fluctuation (denoted by a vertical dashed line). The number $dC$ denotes the position of this transition of behaviour in the number of days elapsed since the beginning of the data set, i.e. 14 March 2020. b Transmission analysis of 11 states [See the explanatory texts in the caption of (a)] c Transmission analysis of 11 states. [See the explanatory texts in the caption of (a)]
Fig. 6 continued
with the nonlinear fitting model \( y = f(x) \). The value of \( Q \) is indicated in each of the scatter plots.

### 4.4.1 Disease transmission and population age

In Fig. 7, 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 the closure of educational institutions, the youth population—primarily the working popula-
The youth population seems to drive the first major transmission. Each column is, respectively, for 10, 20, and 30% asymptomatic transmission rate. These 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.

The urban population also seems to drive the first major transmission. The more is the urban population, the more quickly the day of first major transmission $d_C$ is reached.

In Fig. 8, we show how does the critical transmission rate $\beta_C$ depend on the 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 modelled at 10, 20, and 30% of the primary transmission rate as before.

4.4.2 Disease transmission and urban population

In Fig. 8, we show how does the critical transmission rate $\beta_C$ depend on the 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 modelled at 10, 20, and 30% of the primary transmission rate as before.

4.4.3 The critical day, disease transmission, and economy

In Fig. 9, 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
COVID-19 outbreak in India: an SEIR model-based analysis

Fig. 9 $\beta_C$ vs. $P_{GDP}$: 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 that has driven the transmission.

Fig. 10 $\beta_C$, $\beta_N$ vs. $P_{Econ}$: $P_{Econ}$ is the % of the population below a certain income denoted as the Economy group. The lower is the value of $P_{Econ}$, the richer is the population.

purchasing power), the more is the transmission rate at the first major transmission. So, it is the purchasing power of the population that has also driven the transmission.

Figure 10 contains two rows of which the first row shows the dependence of $\beta_C$ on the economy population $P_{Econ}$ and the second row shows the dependence of the end transmission rate $\beta_N$ (as on 31 August 2020) on $P_{Econ}$. So, the less is the $P_{Econ}$ for a state, the more is the people in the 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. poorer people), lower is the transmission rate at the first major transmission and the longer it requires to have the first major transmission (see Fig. 11). It is interesting to note that this behaviour is in contrast with the behaviour at the later phase of the pandemic. As the pandemic progresses, the transmission rate becomes proportional to
Fig. 11 $d_C$ vs. $P_{Econ}$, $P_{Urban}$: The richer is the population (lower $P_{Econ}$, in the first panel), and the more is the urban population (higher $P_{Urban}$, in the second panel), the quicker it takes to reach the criticality (lower $d_C$).

The $P_{Econ}$, indicating the pandemic fatigue of the low-income group.

As shown in Fig. 11, 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 observations about the affluent and urban people driving the early phase of the pandemic.

5 The overall scenario for the whole country

We now apply our model to the entire country. In what follows, we present the results with the data available from 14 March to 31 August 2020, for a duration of 170 days in Fig. 12. Each row in Fig. 12 has three panels. The first two panels show the results for the cumulative infectives $I$ and daily new infectives $I_{New}$ for the duration of 170 days. The model uses these 170 days of data to model the pandemic. The real data for this period are shown by the thick orange solid line, whereas the data beyond 170 days up to 212 days (for a period of 42 days) are 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. 12, and each row is for the limiting asymptomatic rate $\rho_t = \xi \beta_t$, $\xi = 5, 7.5, 10, 12.5, \text{ and } 15\%$. The mean square error $\epsilon_{\text{MSE}}$, defined as

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

expressed per million of the population for each value of $\xi$ tabulated in Table 6. In the above expression, $Y_i$ is the actual data and $\hat{Y}_i$ is the corresponding estimated value predicted by the model. One can see that the $\epsilon_{\text{MSE}}$ is minimum for $\xi = 7.5\%$. As per the available data till 31 August 2020, the optimum asymptomatic transmission rate of the disease for India was < 10\%. However, as we shall see in the next section, where we use the latest data till March 2021, the suggested asymptomatic transmission rate is $\sim 15\%$.

As can be seen from these figures, the model is able to correctly predict the peak of the pandemic, which occurred around the 180th day (10 September 2020) with daily new infections of about $\sim 10,000$ per day by the end of February 2021. We shall see in the next section that these figures are indeed correct and compare quite well with the logistic curves (see in the next section).

6 Comparison with logistic modelling

In contrast to a generalised compartmental dynamic model of the pandemic of an infectious disease such as the one considered in this paper, it is also quite interesting to model a pandemic with a single logistic function, which can be used to compare the development of the dynamic of the disease in different countries, regions, and areas [4,33,44]. Though a long-term prediction of
Fig. 12  $I$ 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 351 days (28 February 2021). The red coloured dotted parts in the panels show the projection. The number inscribed in each row is the value of $\xi$.

an epidemic based on a logistic function is not useful, it is useful to gain insights into the pandemic as a single function can be used to model the pandemic in diverse circumstances. In this section, we use two logistic functions to model the pandemic based on available data for India (as a country) and compare the results to...
A generalised Gaussian function

\[ I_{\text{new}} = \kappa e^{-\psi(t-t_m)^2}, \quad \psi > 0, \tag{27} \]

which can be used to obtain the logistic function for \( I \) as

\[
I(t) = \kappa_0 + \int I_{\text{new}}(t) \, dt = \kappa_0 + \frac{1}{2} \sqrt{\frac{\pi}{\psi}} \text{erf} \left[ \sqrt{\psi}(t - t_m) \right]. \tag{28}
\]

### 6.3 Comparison

A fitting through nonlinear regression of the above models with the real data is shown in Fig. 13. The regression is carried out with the help of Mathematica through the Levenberg–Marquardt algorithm [30]. The daily new cases \( I_{\text{new}} \) can be found from this model simply by calculating \( I'(t) \), which is also shown in Fig. 13.

In the figure, the comparison of the cumulative infectious \( I \) as predicted from our compartmental model (CM, dashed) with actual data are shown along with the curves as obtained from nonlinear regression with the two logistic models. The number in the bracket in the curve (CM, dashed) indicates the limiting asymptomatic rate constant \( \xi \), as mentioned before. The last panel shows the daily new infections \( I_{\text{new}} \) as obtained from each model plotted with the real data till 14 March 2021. In each panel of Fig. 13, the dotted vertical line indicates the 170th day, which marks the end of real data used for constructing the compartmental model and the rest of the dashed curve is the prediction based on the data used till 170th day. In contrast to this, the curves from the logistic models (GLM and ELM) use all available data. As can be seen from these panels, our prediction based on data till 170th day with \( \xi = 15\% \) almost mimics the real data (the first panel in the third row) and the logistic curves up to about a year (more than six months in advance), which amply proves the efficacy of the model presented here and also establishes its robustness.

This can also be seen from the \( I_{\text{new}} \) curve shown in the second panel of the third row of Fig. 13, where we have shown the \( I_{\text{new}} \) calculated with our model with \( \xi = 15\% \) based on real data only up to 170th day, predicted till 365th day. It also correctly predicts the day with the newest infection as well as the cumulative number of infections.
Fig. 13 Comparison of the infectives \( I \) as predicted from the compartmental model (CM, dashed) with actual data (blue and yellow circles) and the two logistic models. The number in brackets shows the value of \( \xi \). The dotted vertical line marks the 170th day. The last panel shows the daily new infections \( I_{\text{new}} \) as obtained from each model plotted with the real data till March 2021.

It should, however, be noted that India is now seeing the second wave of the infection, which has started from March 2021. The modelling of the second wave will be considerably difficult as by this time the social behaviour of the population is practically largely...
Table 7 Nonlinear regression parameters of the generalised logistic model (GLM)

| End date         | $\rho$ | $\mu$ | $t_{m}$ | $K$       | $\omega$ | $R_0$ |
|------------------|--------|-------|---------|-----------|----------|-------|
| 31 August 2020   | 0.041  | 1.013 | 166     | $6.78 \times 10^6$ | 0.0413   | 1.78  |
| 30 September 2020| 0.050  | 0.430 | 157     | $13.66 \times 10^6$ | 0.0217   | 1.36  |
| 31 October 2020  | 0.039  | 0.863 | 178     | $983 \times 10^6$  | 0.0336   | 1.60  |
| 30 November 2020 | 0.042  | 0.722 | 172     | $10.21 \times 10^6$ | 0.0303   | 1.53  |
| 31 December 2020 | 0.051  | 0.503 | 156     | $10.82 \times 10^6$ | 0.0256   | 1.43  |
| 31 January 2021  | 0.058  | 0.412 | 145     | $11.05 \times 10^6$ | 0.0239   | 1.40  |
| 28 February 2021 | 0.066  | 0.345 | 135     | $11.19 \times 10^6$ | 0.0227   | 1.37  |
| 14 March 2021    | 0.075  | 0.289 | 124     | $11.30 \times 10^6$ | 0.0218   | 1.36  |

Fig. 14 The nonlinear regression curves for the GLM, as more data become available beginning from August 2020 till March 2021. Note how more data points are required to arrive at the present scenario of the pandemic in contrast to the CM, for which the prediction was based on data only up to August 2020

The convergence of nonlinear regression parameters of the GLM model as more and more data become available beyond the 170th day (August 2020) is shown in Table 7 and Fig. 14. (the solid lines). In Table 7, the last two columns are, respectively, the growth rate $\omega$ of the pandemic and the corresponding basic reproduction number $R_0$ as calculated with the GLM during the exponential growth phase of the pandemic which is about the 170th day (see Fig. 12). The basic reproduction number $R_0$ as given by our compartmental model is about $R_0^{CM} \sim 1.2 - 1.8$ (see Fig. 12), which is within the range of the GLM.

7 Summary and conclusions

To summarise, taking into consideration of asymptomatic carriers for the COVID-19 pandemic, an SEIR model for India is proposed with time-varying transmission rates and incubation period, as it is believed uncontrolled with a considerable fraction of the population not sticking to the mandatory COVID protocols.
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 ($\beta_1$) can be written as a fraction of the symptomatic transmission rate ($\beta_2$), the disadvantage of the non-availability of asymptomatic infective data can be incorporated in our model without taking an extra compartment for the asymptomatic population. We have also incorporated a 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 behaviour of the pandemic for a total of 351 days (14 March 2020–28 February 2021). We observed that our predicted results agree well with the data up to 42 days, beyond the end of the data used in the modelling—that was available at the time of undertaking this work. The mean squared error between the predicted data and the actual data is minimum for the asymptomatic transmission rate of 7.5%. However, as of writing the final version of this work, more data became available and it seems that the first phase of the pandemic has passed with its peak by the second week of September 2020 (180 days into the pandemic), as predicted by our model, and has subsided with a daily infection of $\sim 10,000$ cases per day by February 2021, which is also within the range predicted by the model. As such, what we have seen that the prediction of our model based on data up to the 170th day into the pandemic agree quite well with the real data till February 2021, about the 351th day into the pandemic. However, by March 2021, the second surge of infections has started and now we are witnessing the second phase of the pandemic. Apparently, at present there are no enough data to analyse the second phase and would possibly require different sets of conditions as the second phase is believed to be different from its first phase in terms of behaviour and is beyond the scope of the present work. In the last section, we have also analysed the latest data with the help of two logistic models, and the logistic curves agree very well with the prediction of our model with an asymptomatic transmission rate of 15%.

We also proposed a way to pinpoint the time when the first major transmission occurs (we called it a critical day) for a particular state, and the 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 states with higher youth population have larger transmission rates on the critical day. If a state has a higher child population, the critical transmission rate has been found to be lower for that state. We also find that the higher is 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) experience a higher rate of transmission at the critical day. These are some interesting results of this work that reflect the correlation between the behaviour of the pandemic with the 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 progresses, it has been gradually found to be driven by the people with relatively lower income, demonstrating the so-called pandemic fatigue. On the basis of the available data till the 170th day for India, we have found an asymptomatic transmission rate of $\sim 7.5\%$ to be an optimum rate for the best-fit parameters. However, a projection predicted by our model with an asymptomatic transmission rate of 15% closely agrees with the latest data that are available till writing of this paper, which is also suggested by the logistic fittings. As is shown in Fig. 12, the value of $R_0$ should have fallen to $\sim 0.5$ by the end of April 2021, when the disease could have been taken as contained, if not for the second surge of infections, which India is presently witnessing. Our equilibria analysis shows that in India, an endemic equilibrium can reach when the infectives can reach about 415 million, which is very far from the cumulative total of about 11.4 million cases (as of mid-March 2021). Thus, we can probably safely assume that India will not ever reach an EE state, now that the vaccination program has already started.

To conclude, we have proposed a compartmental model for modelling the COVID-19 pandemic in India, which shows excellent agreement with the real data well beyond the intended prediction period. We have also shown how the pandemic can be related to the socio-economic status and demography of a particular region of India. The results of this model can be
used for early dynamic preventive measures for similar situations in 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).

**Availability of data** All data analysed in this study are publicly available in the Gauhati University Data Repository—‘COVID-19 Data (India 2020–21)’ at https://web.gauhati.ac.in/data/covid-19.

**Declarations**

**Conflicts of interest** The authors declare that they have no conflict of interest.

**References**

1. Analysis of state budgets (2020). https://www.prsindia.org/parliamenttrack/budgets/state(2020)
2. 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)
3. Aviv-Sharon, E., Aharoni, A.: Generalized logistic growth modeling of the COVID-19 pandemic in Asia. Infect. Dis. Model. 5, 502–509 (2020)
4. 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)
5. Barman, M.P., Rahman, T., Bora, K., Borgohain, C.: Covid-19 pandemic and its recovery time of patients in India: a pilot study. Diabetes Metab. Syndr. 14, 1205–1211 (2020)
6. Census of India 2011 (2011). https://censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf
7. Countries in the world by population (2020). https://www.worldometers.info/world-population/population-by-country
8. COVID-19 India Org Data Operations Group (2020). https://api.covid19india.org
9. Fanelli, D., Piazza, F.: Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 134, 109761 (2020)
10. Fernandez, B.: Poor practices: contestations around “below poverty line” status in India. Third World Q. 31(3), 415–430 (2010)
11. Giordano, G., et al.: Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. 26, 855–860 (2020)
12. He, S., Peng, Y., Sun, K.: SEIR modeling of the COVID-19 and its dynamics. Nonlinear Dyn. 101, 1667–1680 (2020)
13. Ibrahim, M.A., Al-Najafi, A.: Modeling, control, and prediction of the spread of COVID-19 using compartmental, logistic, and gauss models: a case study in Iraq and Egypt. Processes 8(11), 1400 (2020)
14. John Hopkins University, Coronavirus Resource Center (2020). https://coronavirus.jhu.edu
15. Kakar, A., Nundy, S.: Covid-19 in India. J. R. Soc. Med. 113(6), 232–233 (2020)
16. Kaxiras, E., Neofotistos, G., Angelaki, E.: The first 100 days: modeling the evolution of the COVID-19 pandemic. Chaos Solitons Fractals 138, 110114 (2020)
17. Li, M.Y.: An introduction to mathematical modeling of infectious diseases. Springer, New York (2018)
18. 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)
19. 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)
20. Liu, X., Zheng, X., Balachandran, B.: Covid-19: data-driven dynamics, statistical and distributed delay models, and observations. Nonlinear Dyn. 101, 1527–1543 (2020)
21. 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)
22. Martcheva, M.: An introduction to mathematical epidemiology. Springer, New York (2015)
23. Miller, E.: The Vaccine Book, chap. 1, pp. 37–50. Academic Press, Cambridge (2003)
24. Ministry of Health and Family Welfare (2020). https://www.mohfw.gov.in
25. Nabi, K.N.: Forecasting COVID-19 pandemic: a data-driven analysis. Chaos Solitons Fractals 139, 110046 (2020)
26. National Accounts Data (2020). http://mospi.nic.in/download-tables-data
27. Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. 7(4), 308–313 (1965)
28. Pal, S.C., et al.: Threats of unplanned movement of migrant workers for sudden spurt of COVID-19 pandemic in India. Cities (2020)
29. Population composition, chapter 2 (2011). https://censusindia.gov.in/vital_statistics/9Chap%202%20-%202011.pdf
30. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes. Cambridge University Press, Cambridge (2007)
31. Provisional Population Totals, Census of India 2011, vol. 1 (2011). https://censusindia.gov.in/2011-prov-results/paper2/data_files/india/paper2_1.pdf
32. Quadri, S.A.: COVID-19 and religious congregations: implications for spread of novel pathogens. Int. J. Infect. Dis. 96, 219–221 (2020)
33. Richards, F.J.: A flexible growth function for empirical use. J. Exp. Bot. 10(29), 290–300 (1959)
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. Sample Registration System, Baseline Survey 2014 (2014). https://censusindia.gov.in/vital_statistics/BASELINE%20TABLES07062016.pdf
36. 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)
37. Sarkar, K., Khajanchi, S., Nieto, J.J.: Modeling and forecasting the COVID-19 pandemic in India. Chaos Solitons Fractals 139, 110049 (2020)
38. SDG India index, NITI Aayog (2020). https://sdgindiaindex.niti.gov.in/#/ranking
39. Sengupta, S., Jha, M.K.: Social policy, COVID-19 and impoverished migrants: challenges and prospects in locked down India. Int. J. Commun. Soc. Develop. 2(2), 152–172 (2020)
40. State/UT wise Adhaar Saturation (2020). https://uidai.gov.in/images/state-wise-aadhaar-saturation.pdf
41. 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)
42. Wu, K., Darcet, D., Wang, Q., Sornette, D.: Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. Nonlinear Dyn. 101, 1561–1581 (2020)
43. Young, P.: Everything you wanted to know about data analysis and fitting but were afraid to ask, 1st edn. Springer, New York (2015)
44. 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)

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