Spatio-temporal predictive modeling framework for infectious disease spread

Sashikumaar Ganesan[1] **and Deepak Subramani**[1]

[1]Department of Computational and Data Sciences, IISc Bangalore
sashi@iisc.ac.in, deepakns@iisc.ac.in

ABSTRACT

A novel predictive modeling framework for the spread of infectious diseases using high dimensional partial differential equations is developed and implemented. A scalar function representing the infected population is defined on a high-dimensional space and its evolution over all directions is described by a population balance equation (PBE). New infections are introduced among the susceptible population from non-quarantined infected population based on their interaction, adherence to distancing norms, hygiene levels and any other societal interventions. Moreover, recovery, death, immunity and all aforementioned parameters are modeled on the high-dimensional space. To epitomize the capabilities and features of the above framework, prognostic estimates of Covid-19 spread using a six-dimensional (time, 2D space, infection severity, duration of infection, and population age) PBE is presented. Further, scenario analysis for different policy interventions and population behavior is presented, throwing more insights into the spatio-temporal spread of infections across disease age, intensity and age of population. These insights could be used for science-informed policy planning.

Introduction

Epidemic modeling and forecasting has gained renewed interest since the late 2019 when the world was affected by the novel coronavirus pandemic (named Covid-19). Several computational studies to predict the human-to-human spread of Covid-19 have been reported[1–4]. Most of these efforts have been based on compartmental models and stochastic models (including agent-based models)[5]. In compartmental models (e.g., SIR, SEIR, SEIRS, DELPHI[6] etc), the population is divided into different compartments and the dynamics of the different compartments are modeled by a system of coupled ordinary differential equations (ODE). Here, the interaction among compartments is usually deterministic, whereas random processes are used to the spread of infections in stochastic models. Agent-based models are stochastic models that undertake a bottom-up approach of modeling individual members of a population and the dynamics of their interaction in terms of probabilities of movement and contact.

Furthermore, for science-informed policy intervention and public health planning, more than the total number of infections, it is essential to have more insightful predictions, e.g., infected population distribution across their age and level of infection severity. The population distribution across the age of infection is crucial for planing antiviral treatments, quarantine, ventilator support and contact tracing. This requirement necessitates a comprehensive and computationally efficient predictive modeling framework. Even though these features could be incorporated in ODE-based compartmental and stochastic agent-based models, it is very complex and computationally expensive. To overcome these challenges, we propose a novel partial differential equation based spatio-temporal predictive modeling framework for forecasting the spread of infectious disease in heterogeneous populations in open geographies. The roots for our model lie in the population balance equations that are popular in chemical engineering and process studies[7].

In our proposed model, the infected population density is defined as a scalar field on a high-dimensional space. Specifically for predicting the spread of Covid-19, a six-dimensional model is presented. The first three dimensions are the space and time, and the other three are the infection severity, duration of the infection (i.e., time since infection), and age of the population. New infections, impact of quarantine, testing, contact tracing, immunity, intervention policy impact, health infrastructure, recovery, and death are all modeled on this six-dimensional space based on data-driven functions (where available), and/or simple algebraic and integral functions. Notably, our PDE-based model in the present paper is more compact and a versatile description of the spread of the disease compared to compartmental models, and computationally efficient compared to agent-based models. To showcase the capability of our distribution-based predictive modeling framework for infectious disease spread, we apply it to model and predict the spread of Covid-19 in India. Further, we present a scenario analysis, which could be used to draw insights for policy interventions.

Results

The Population Balance Model

Let T_∞ be a given final time and $\Omega := \Omega_x \otimes \Omega_\ell$ be the computational domain of interest. Here, $\Omega_x \subset \mathbb{R}^2$ is the spatial domain defining the geographical region of interest and $\Omega_\ell \subset \mathbb{R}^n$, where n is the number of internal directions. Each of the n-internal directions represents the property of the population on which a distribution needs to be predicted. Suppose the properties of interest are the infection severity, duration of the infection and age of population, then a model with three internal directions could be used as follows. Let $\Omega_\ell := L_v \times L_d \times L_a$ be the internal domain, where $L_v = [0,1]$ denotes the infection severity interval, $L_d = [0, d_\infty]$ denotes the duration of infection, d_∞ is the maximum duration of infection, $L_a = [0, a_\infty]$ denotes the age interval and a_∞ is the maximum age of the population. The infection index $\ell_v \in L_v$ quantifies the severity of the population being infected by the disease. Specifically, the population with infection index $\ell_v = 0$ is completely disease free, with $\ell_v = 1$ has maximum severity, with $\ell_v \geq v_{\text{sym}}$ shows symptoms and those with $\ell_v < v_{\text{sym}}$ are asymptomatic. The duration of infection index $\ell_d \in L_d$ quantifies the time since a population has been exposed to and contracted the disease. Specifically, the population that just contracted the disease has $\ell_d = 0$. Typically, a person is asymptomatic until they reach $\ell_v = v_{\text{sym}}$, and the duration elapsed ℓ_d is the incubation period in which the disease is sub-clinical and that population is actively spreading the disease. After recovery, a population doesn't necessarily go to $\ell_v = 0$, rather they reach $\ell_v < v_{reco}$.

Let $I(t,x,\ell_v,\ell_d,\ell_a)$ be the infected number density function of the population. To describe the evolution of the active infected population size distribution, we propose the population balance equation

$$\begin{aligned}
\frac{\partial I}{\partial t} + \nabla_x \cdot (\mathbf{u} I) + \nabla_\ell \cdot (\mathbf{G} I) + CI &= F && \text{in} \quad (0,T_\infty] \times \Omega_x \times \Omega_\ell\,, \\
I(t,x,\ell) &= g_n && \text{in} \quad (0,T_\infty] \times \partial\Omega_x^- \times L_v \times L_d \times L_a\,, \\
I(t,\mathbf{x},\ell_v,0,\ell_a) &= B_{\text{nuc}} && \text{in} \quad (0,T_\infty] \times \Omega_x \times L_v \times L_a\,, \\
I(t,\mathbf{x},0,\ell_d>0,\ell_a) &= 0 && \text{in} \quad (0,T_\infty] \times \Omega_x \times L_d \times L_a\,, \\
I(0,x,\ell) &= I_0 && \text{in} \quad \Omega_x \times \Omega_\ell\,.
\end{aligned} \tag{1}$$

Here, $\mathbf{u}$ denotes the advection vector that quantifies the multiscale spatial movement of the population in a differential neighbourhood of Ω_x (e.g., migrant laborers, daily commute for work, logistics-related travel, periodic gathering for religious and social events) , $\mathbf{n}$ is the outward unit normal vector to Ω_x, $\partial\Omega^- := \{x \in \partial\Omega_x \mid \mathbf{u}\cdot\mathbf{n} < \mathbf{0}\}$, g_n is the flux that quantifies the net addition of an infected population into Ω_x from outside (the spatial movement of the population across the border of the domain $\partial\Omega_x$), and I_0 is the initial distribution of infected population. Further, $\mathbf{G} = (G_{\ell_v}, G_{\ell_d}, G_{\ell_a})^T$ is the internal growth vector, where

$$G_{\ell_v} = \frac{d\ell_v}{dt} = G_{\ell_v}(\ell_a, \beta, \gamma(\ell_a), \alpha(x)), \qquad G_{\ell_d} = \frac{d\ell_d}{dt} = 1, \qquad G_{\ell_a} = \frac{d\ell_a}{dt} = 1. \tag{2}$$

Here, β is the immunity of the infected population, γ is the pre-medical history of the infected population and α is the effective treatment index. Next, we define the rate term $C = C_R + C_{ID}$, where $C_R(t,\mathbf{x},\ell_v,\ell_d,\ell_a)$ is a recovery rate function that quantifies the rate of recovery of the population from the infection, and $C_{ID}(t,\mathbf{x},\ell_v,\ell_d,\ell_a)$ is the infectious death rate. We also define a source term $F = C_T(t,\mathbf{x},\ell_v,\ell_d,\ell_a)$ that quantifies the point-to-point movement of infected population (e.g., by air, train etc) within Ω_x, which are not included in $\mathbf{u}$ and g_n. Moreover, C_T and $\mathbf{u}$ need to be defined in such a way that the net internal movement of infected population within Ω_x is conserved. Moreover, B_{nuc} is the nucleation function that quantifies the infection transmission from the infected to the susceptible population and it is a function of several parameters as follows

$$B_{\text{nuc}} = B_{\text{nuc}}\,(t, \Omega, \sigma, H, S_D, N_S, N_Q, I)\,. \tag{3}$$

Here, σ, H and S_D are the interactivity, hygiene and social distancing indices respectively. Finally, the total population $N(t)$ at a given time $t \in (0, T_\infty]$ is defined by

$$N(t) = N_S(t) + N_R(t) + N_I(t) + N_Q(t) - N_{ID}(t) + N_B(t) - N_D(t), \quad N_Q(t) = \int_\Omega \gamma_Q(t,\mathbf{x},\ell_v,\ell_d,\ell_a) I(t,\mathbf{x},\ell_v,\ell_d,\ell_a)\,dx\,d\ell\,,$$

$$N_I(t) = \int_\Omega I(t,\mathbf{x},\ell_v,\ell_d,\ell_a)\,d\Omega, \quad N_R(t) = \int_\Omega C_R I(t,\mathbf{x},\ell_v,\ell_d,\ell_a)\,d\Omega\,, \quad N_{ID}(t) = \int_\Omega C_{ID} I(t,\mathbf{x},\ell_v,\ell_d,\ell_a)\,d\Omega\,.$$

Here, N_S, N_B, N_R, N_I, N_Q N_{ID} and N_D are the number of susceptible, newborn, recovered, infected (symptomatic/asymptomatic), quarantined, infectious death and natural death populations, respectively. The given initial and boundary conditions and the above defined parameters close the population balance system.

Modeling of Parameters

The proposed population balance model (1) is comprehensive and built on the basis of several parameters as defined above. In this section, we describe the modeling of each parameter.

Nucleation Term

The nucleation term B_{nuc} quantifies the new infection number density that is added to the system at $\ell_d = 0$ for all ordinates ℓ_v, ℓ_a, $\mathbf{x}$ and t. Depending on how the susceptible population interacts with the infected population, new infections are added to the system. We call this addition as *nucleation* (borrowing terminology from process studies), which is modelled as

$$B_{\text{nuc}} = R(t,\mathbf{x},\ell_v,\ell_a)\int_{\Omega_\ell}[1-\gamma_Q(t,\mathbf{x},\ell_v,\ell_d,\ell_a)]I(t,\mathbf{x},\ell_a,\ell_d,\ell_v)\mathrm{d}\ell_v\mathrm{d}\ell_d\mathrm{d}\ell_a\,, \tag{4}$$

$$R(t,\mathbf{x},\ell_v,\ell_a) = R_0(t,\mathbf{x})f_1(t,\sigma)f_2(t,H)f_3(t,S_D)f_4(t,\mathbf{x},\ell_a)f_5(\ell_v)\,, \tag{5}$$

$$\gamma_Q(t,\mathbf{x},\ell_v,\ell_d,\ell_a) = \frac{1}{1+\exp\left(-(\ell_v-v_{\text{sym}})/b_v\right)}\frac{1}{1+\exp\left(-(\ell_d-d_{\text{sym}})/b_d\right)}\frac{1}{1+\exp\left(-(\ell_a-a_{risk})/b_a\right)}. \tag{6}$$

Here, $\gamma_Q \in [0,1]$ in (4) determines the fraction of infected population in quarantine and it can be modeled as in equation (6). Further, the factor γ_Q is dependent on the level of screening including testing, strictness of enforcing isolation and compliance of susceptible general public. Suppose $\gamma_Q = 1$, i.e., if the entire infected population is kept under strict isolation, newly infected population will be zero and eventually there will be no spread of disease. However, due to economic, social and democratic reasons, implementing such a strategy is nearly impossible and there is bound to be spread, i.e., $\gamma_Q < 1$. Moreover, the integral on the right hand side of equation (4) is the total non-quarantined number density of the infected population at $(t,\mathbf{x})$, and R is the rate at which the non-quarantined population infects the susceptible population. The factor R is modelled as in equation (5), where R_0 is the basic reproduction rate,

$$\begin{aligned}
f_1(t,\sigma) &= \left[\tfrac{1}{1+\exp(-(\sigma(t)-\sigma_c)/b_\sigma)}\right], & f_2(t,H) &= \left[1-\tfrac{1}{1+\exp(-(H(t)-H_c)/b_H)}\right]\\
f_3(t,S_D) &= \left[1-\tfrac{1}{1+\exp\left(-(S_D(t)-S_{D_c})/b_{S_D}\right)}\right], & & \\
f_4(t,\mathbf{x},\ell_a) &= a_4\exp\left(-\tfrac{(\ell_a-b_4)^2}{c_4^2}\right), & f_5(\ell_v) &= \begin{cases} 3\sqrt{\frac{2}{\pi}}\exp\left(\frac{-(\ell_v-v_{\text{sym}})^2}{2(v_{\text{sym}}/3)^2}\right) & 0\le v<v_{\text{sym}}\,,\\ 3\sqrt{\frac{2}{\pi}}\exp\left(\frac{-(\ell_v-v_{\text{sym}})^2}{2((1-v_{\text{sym}})/3)^2}\right) & v_{\text{sym}}\le v\le 1.\end{cases}
\end{aligned}$$

Here, the interactivity index $\sigma \in [0,1]$, hygiene index $H \in [0,1]$, and social distancing index $S_D \in [0,1]$. Suppose $\sigma = 0$ then everything is under perfect lockdown and $R \to 0$. In case $S_D = 1$, everyone is following perfect social distancing and $R \to 0$. Moreover, the newly infected population has to be added at different age (ℓ_a) and infection (ℓ_v) levels for which the factors f_4 and f_5 are introduced. We propose to use logistic functions fitted to data from literature for f_1, f_2, f_3,; the normalized demography at $(t,\mathbf{x})$ for $f_4(t,\mathbf{x},\ell_a)$, and a Gaussian mixture with two components so that maxima is at v_{sym} and tails are proportional to the interval length over $[0,v_{\text{sym}}]$ and $[v_{\text{sym}},1]$ for $f_5(\ell_v)$. In addition, the constant in f_5 is chosen such that the integral of f_5 over its support is one. This condition is imposed to ensure that R_0 can be interpreted as the basic reproduction rate used in standard epidemiological models[8]. The parameters v_{sym}, d_{sym}, a_{risk}, b_v, b_d, b_a, b_σ, b_H, b_{S_D}, σ_c, H_c, S_{D_c} can be estimated from experimental and clinical evidence. Furthermore, in light of new evidence, the functional forms of f_1 to f_5 can be easily modified. Finally, $\sigma(t)$, $S_D(t)$ and $H(t)$ change over time due to increased awareness, government measures and compliance by people.

Growth Factor

The growth factor G_{ℓ_v} quantifies how the infected number density is advected along the direction of l_v, that is, how the infection becomes mild to severe or critical in the infected population. We can model it as function of the medical history, immunity of the population, which in turn are functions of the age l_a, treatment and socio-economic status. Nevertheless, as a simple first order model, we model it as a nonlinear function of the age, i.e.,

$$G_{\ell_v}(\ell_a) = K_g(\ell_a - a_{risk})^p\,, \tag{7}$$

where K_g is a non-dimensionalization factor, p is a power of nonlinearity and l_{a_c} is the age offset.

Recovery Rate and Infectious Death Rate

In general, the recovery rate C_R and infectious death rate C_{ID} depends on ℓ_v, and in turn are functions of hospital facilities, age, and health state of the population. These rates can be modeled directly from historical data for all ordinates ℓ_v, ℓ_d, ℓ_a. For the exact functional forms refer to the Appendix.

Initial Infection Number Density

The initial number density $I_0(\mathbf{x},\ell_v,\ell_d,\ell_a)$ can be estimated directly from available official data at the day of starting the simulation. However, the available data is only in-terms of total number of tested and confirmed cases at a $\mathbf{x}$-location and the

dependence on $(\ell_v,\ \ell_d,\ \ell_a)$ needs to be estimated via appropriate data-driven and analytical functions. As such, first we utilize data from a period of 14 days, along with the log-normal distribution of incubation period[9] to calculate the initial number density $N_D(\mathbf{x})$ at all the spatial points $\mathbf{x}$, but integrated over the three internal ordinates (ℓ_v, ℓ_a, ℓ_d), i.e.,

$$N_D(\mathbf{x}) = \sum_{i=1}^{i=14} N_i \frac{1}{i a_2 \sqrt{2\pi}} \exp\left(-\frac{(\log i - b_2)^2}{2a_2^2}\right), \tag{8}$$

where $N_i(\mathbf{x})$ is the data of tested and positive. For the distribution along the internal ordinates, we propose to use the following initial infection number density distribution

$$I_0(\mathbf{x}, \ell_v, \ell_d, \ell_a) = N_D(\mathbf{x}) \left[a_1 \exp\left(-\frac{(\ell_a - b_1)^2}{c_1^2}\right)\right] \left[\frac{1}{\ell_d a_2 \sqrt{2\pi}} \exp\left(-\frac{(\log(\ell_d) - b_2)^2}{2a_2^2}\right)\right] [f_5(\ell_v)]. \tag{9}$$

Here, the first term in the square brackets is the normalized demography function (same as f_4), second term is the log-normal incubation period function with fitted[9] $a_2 = 0.42$ and $b_2 = 1.62$, and f_5 is same as before.

Covid-19 Epidemic Spread Predictions

To exhibit the capabilities of the proposed model, the forecast of Covid-19 spread in India is presented here. The numerical scheme and the fitted model parameters are given in the Appendix. The proposed model and numerical schemes are implemented in our in-house finite element package[10,11] and have been verified in our earlier studies with applications to process engineering[12,13].

With the spread of Covid-19 in India, the federal government imposed a nation-wide lockdown from March 25, 2020. To simulate the spread of infections starting from March 23, 2020, the initial distribution of infected population is estimated using the data of active cases from March 23 to April 5 according to equation 8 and equation 9. Then the infection spread forecast for one year is computed by solving the PBE system (equation 1). Further, data until June 21, 2020 is utilized to select the parameters (e.g., S_D, C_R, C_{ID}, γ_Q) that best explains the actual data. Thereafter, the control parameter S_D is varied to perform scenario analyses as presented next.

Scenario Analysis

Different future scenarios are predicted by varying $S_D(t)$ based on anticipated individual behavior (social distancing, hygiene practice, compliance to government rules etc.) and government policies (quarantine rules, lockdown rules etc.). The first scenario, named *Current Trend* follows business as usual assuming further relaxation to lockdown rules. A second variant named *Better Scenario* assumes better compliance in the social distancing and other measures to control the spread of the disease. Sunday, and Sunday & Wednesday lockdowns are imposed on the *Current Trend* scenario to formulate the third and fourth scenarios respectively. These lockdown scenarios are introduced to measure the impact of periodic lockdowns on the effectiveness of these strategies to control the disease spread. The active ($N_I(t)$), recovered (cumulative N_R), deaths (cumulative

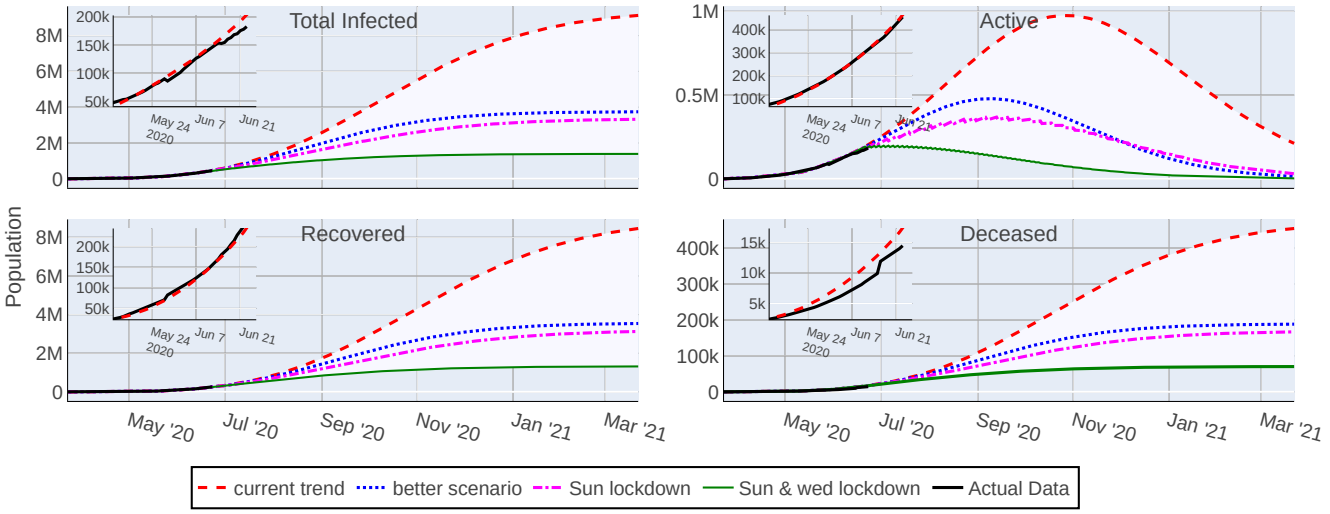

Figure 1. Time series forecast of active, total infections, recovered and deceased cases of Covid-19 in India from Mar 23, 2020 to Mar 22, 2021. The inset shows a zoom with comparison of the model forecast with the data until June 21, 2020

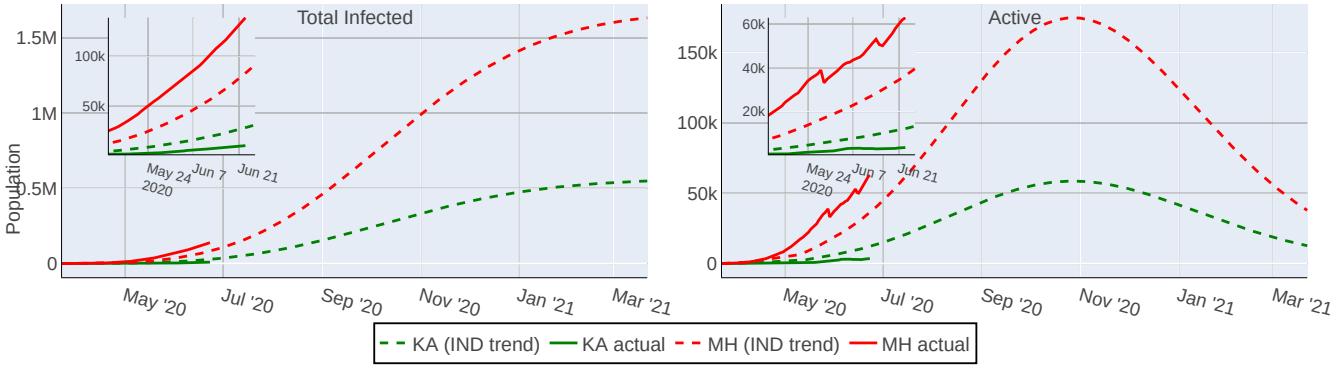

Figure 2. The actual data and the predictions computed with the national trend based parameters for the states of Maharashtra and Karnataka in India.

N_{ID}) and total (sum of active, recovered and deaths) predicted by the four scenarios for the duration between March 23, 2020 and March 22, 2021 are shown as time-series plots in Fig. 1. In the *Current Trend*, a peak of 0.975 million 'Active Cases' is hit in the last week of October 2020, and there will be around 21 million 'Active Cases', 450,000 deaths and 9.1 million total cases at the end of March 2021. In the *Better Scenario*, a lower peak (compared to the current trend) of 0.478 million 'Active Cases' is hit in the second week of September 2020, and there will be around 14,200 'Active Cases', 0.188 million deaths and 3.74 million total cases at the end of March 2021. The weekly lockdown scenarios assume that a complete lockdown is imposed on Sunday or Sunday and Wednesday. During this lockdown, there is a complete restriction of people's movement similar to the nationwide lockdown imposed between Mar 25 and April 14 in India. With Sunday Lockdown, a peak of 0.365 million 'Active Cases' is sustained for about two weeks during 5-20 September 2020, and there will be around 30,200 thousand 'Active Cases', 0.167 million deaths and 3.32 million total cases at the end of March 2021. With Sunday and Wednesday lockdown, a peak of 0.197 million 'Active Cases' is sustained for the period 27 June to 15 July 2020, and there will around 2,800 'Active Cases', 70,300 deaths and 1.39 million total cases at the end of March 2021. The insets in each panel of Fig. 1 show the comparison with actual data and thereby validate the model. In addition, the time series plots for other scenarios including a worse-case scenario can be found at IISc-Model website[14].

In order to compare the performance of all states in India with the national trend, a uniform set of parameters is used for state-wide computations. In particular, the parameters are fitted by minimizing the error between the national data and the sum of the respective state predictions. Fig. 2 shows the actual data and the computed distribution using the above set of parameters for the states of Maharashtra and Karnataka. We can see that Karnataka has done better, whereas Maharashtra has done worse compared to the national trend. These insights can be used by the authorities to introduce state-wise lockdown policies and to plan infrastructure for quarantine, treatments etc. The performance of other states can be seen at IISc-Model website[14].

Population distribution

Our PBE model in fact predicts the distribution of the infected population over all the internal ordinates ℓ_v, ℓ_d, ℓ_a. In the previous section, we have shown only the total number of infected, recovered and deceased populations. Now, to showcase one of the unique features of the model, we present and discuss the predicted population distribution for the Sunday lockdown scenario (Table 1). Figure 3 shows the predicted distribution of active Covid-19 infected population over the ordinates ℓ_v, ℓ_d and ℓ_a at different time instances, that is, on day 60, 120, 180, 240, 300 and 365 respectively.

Predicting the severity of the infected population is crucial to plan the hospital requirements including antiviral treatments, quarantine, hospitalisation, ventilator support, and oxygen support. In particular, the information of asymptomatic and symptomatic infected population helps the policymakers to plan quarantine rules. Moreover, the death rate is a function of the severity of infection and is crucial to predict the causalities arising from the infection spread. The infectious period or the age of infection plays a key role in epidemic modeling. Classical models usually assume a constant infection period. However, the recovery and the death of the patient depend on the immunity, age and health of the patient, medical treatments etc and thus the infectious period need not be a constant. The age of infection is considered as an independent variable in the proposed model. Figure 4 depicts the distribution of recovered population across different parameters on day 60, 120, 180, 240, 300 and 365, respectively. Moreover, the age of infection is key to model the recovery rate, quarantine, etc. Further, the prediction with age of infection, especially at initial stages, is key to plan for testing and to make effective decisions on quarantine, hospitalization and discharging from hospitals. Finally, the age of the population is pivotal in epidemics like Covid-19 since it affects the infants and elders severely. It is incorporated into the proposed model and in fact the newly infected population is distributed across the age of the susceptible population. Moreover, the response to the antiviral treatment, death and recovery rates depend on age-specific health complications such as diabetics, cardiovascular disease can also be incorporated in the proposed model.

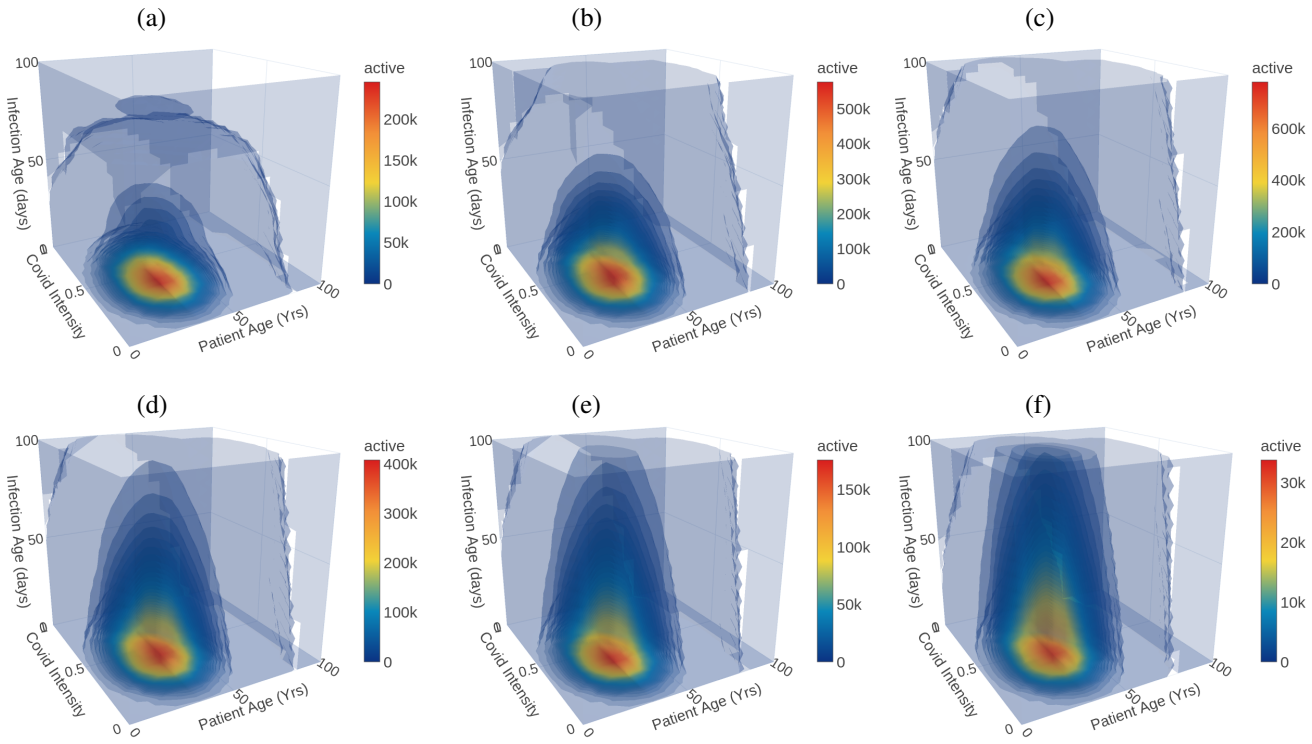

Figure 3. Distributions of Covid-19 population at different time instances, (a) t=60, (b) t=120, (c) t=180, (d) t=240, (e) t=300 and (f) t=365.

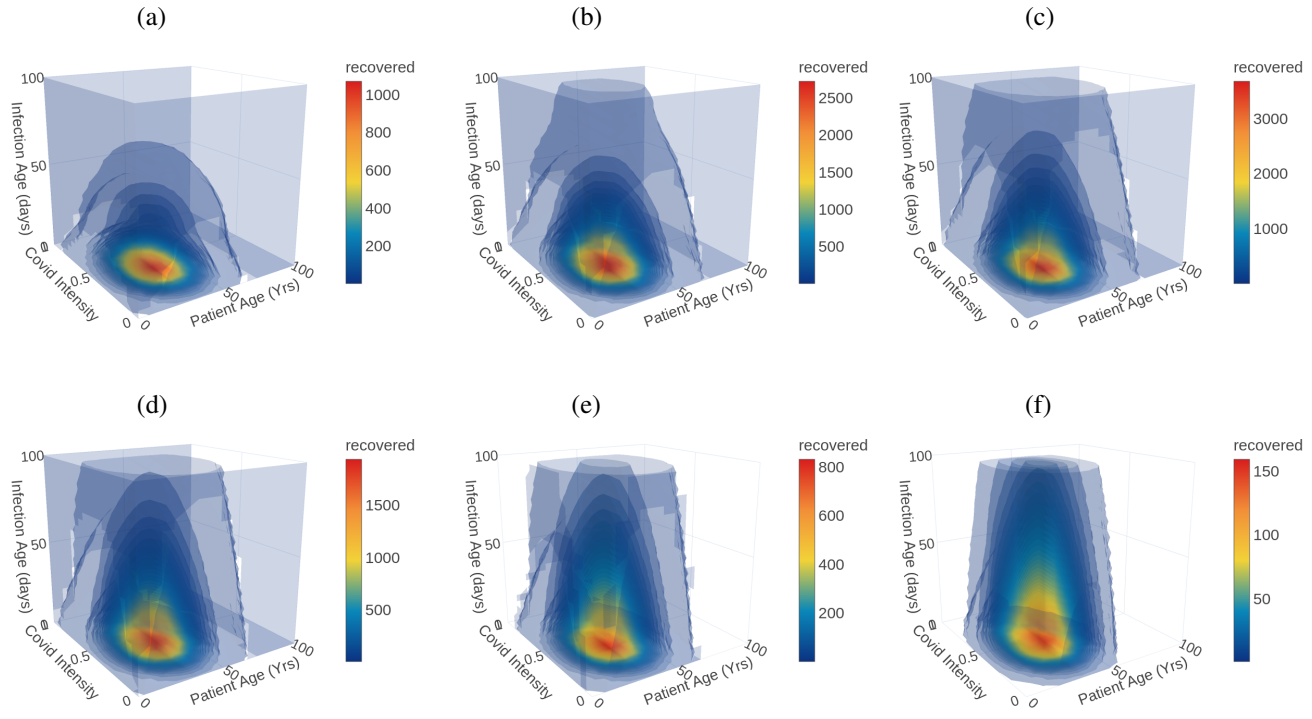

Figure 4. Distributions of recovered population from Covid-19 at different time instances, (a) t=60, (b) t=120, (c) t=180, (d) t=240, (e) t=300 and (f) t=365.

Discussion

Our spatio-temporal modeling framework is the first comprehensive partial differential equation model for predicting infectious disease spread. Computationally, our model is efficient compared to agent-based stochastic models. Mathematically, our PDE system is more compact and comprehensive compared to ODE-based compartmental models. Specifically, the PDE is a continuum description whereas the compartmental models are a discrete representation of the population. Crucially, in contrast to the existing models, our model provides an insight into the distribution of infected population (presented in previous sections). This information is important to plan policy interventions, especially in Covid-19 like pandemics. Not only prognostic estimates, but also diagnostic estimates for more detailed analysis using distribution can be performed with the proposed framework.

With more data and employing data-driven and machine learning approaches, we could further refine the parameters and functional forms of different model components to derive more insightful predictions. For example, to derive insights into the reopening of workplace and educational institutions, the nucleation and advection vector could be modeled to account for interactions between different age groups and movement of people between homes and these places. The potential options for refining the model are virtually endless. In particular, there is no restriction on the choice of number of internal coordinates. For example, in addition to ℓ_v, ℓ_d, ℓ_a, profession, mobility history, etc can also be added as internal coordinates.

Even though we have emphasized Covid-19 pandemic in the present paper, the proposed model can readily be used for forecasting any other infectious disease spread. In future, a data assimilative framework for real-time update of forecasts can also be implemented.

References

1. Singh, R. & Adhikari, R. Age-structured impact of social distancing on the covid-19 epidemic in india. *arXiv preprint arXiv:2003.12055* (2020).
2. Harsha, P. *et al.* Covid-19 epidemic study ii: Phased emergence from the lockdown in mumbai. *arXiv preprint arXiv:2006.03375* (2020).
3. Pandey, G., Chaudhary, P., Gupta, R. & Pal, S. Seir and regression model based covid-19 outbreak predictions in india. *arXiv preprint arXiv:2004.00958* (2020).
4. Ranjan, R. Predictions for covid-19 outbreak in india using epidemiological models. *medRxiv* (2020).
5. Vynnycky, E. & White, R. *An introduction to infectious disease modelling* (OUP oxford, 2010).
6. Bertsimas, el. al. *Covid Analaytics Website* (2020: accessed June 10, 2020 at https://www.covidanalytics.io).
7. Ramkrishna, D. & Mahoney, A. W. Population balance modeling. promise for the future. *Chem. Eng. Sci.* **57**, 595–606 (2002).
8. Grassly, N. C. & Fraser, C. Mathematical models of infectious disease transmission. *Nat. Rev. Microbiol.* **6**, 477–487 (2008).
9. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals internal medicine* **172**, 577–582 (2020).
10. Ganesan, S. *et al.* An object oriented parallel finite element scheme for computations of pdes: Design and implementation. *2016 IEEE 23rd Int. Conf. on High Perform. Comput. Work. (HiPCW)* 2–11, DOI: 10.1109/HiPCW.2016.023 (2016).
11. Wilbrandt, U. *et al.* Parmoon – a modernized program package based on mapped finite elements. *Comput. Math. with Appl.* **74**, 74–88, DOI: 10.1016/j.camwa.2016.12.020 (2016).
12. Ganesan, S. & Tobiska, L. Operator-splitting finite element algorithms for computations of high-dimensional parabolic problems. *Appl. Math. Comp.* **219**, 6182–6196 (2013).
13. Ganesan, S. An operator-splitting Galerkin/SUPG finite element method for population balance equations: Stability and convergence. *ESAIM: M2AN* **46**, 1447–1465 (2012).
14. Ganesan, S. & Subramani, D. *IISc-Model Website* (2020: accessed May 3, 2020 at https://cmg.cds.iisc.ac.in/covid/).
15. Ganesan, S. & Tobiska, L. *Finite Elements: Theory and Algorithms*. Cambridge IISc Series (Cambridge University Press, 2016).
16. Ganesan, S. & Tobiska, L. An operator-splitting finite element method for the efficient parallel solution of multidimensional population balance systems. *Chem. Eng. Sci.* **69**, 59–68 (2012).
17. Crowdsourced. *Covid 19 India* (2020: accessed April 10, 2020 at https://www.covid19india.org/).
18. Statista. *Statista Website* (2020: accessed April 10, 2020 at https://www.statista.com/statistics/1110522/india-number-of-coronavirus-cases-by-age-group/).

Acknowledgements

DS acknowledges the partial support from IISc Start-up grant, DST-INSPIRE Faculty Research Grant (04/2018/003591), and Arcot Ramachandran Young Investigator Award. SG acknowledges the support from SERB and DRDO for the grants that supported development of ParMooN.

Author contributions statement

Both authors contributed equally and reviewed the manuscript.

Appendix

As the focus of the present paper is to introduce the spatio-temporal predictive modeling framework for infectious disease spread, we use simple estimates and algebraic relations for certain parameters. Further, a few more assumptions are made on some parameters due to lack of actual data. Nevertheless, the features of the model and the insights into the prediction of the spread can be seen even with these assumptions.

Numerical Scheme

Let $T_\infty = 365$ days, $d_\infty = 400$ days, $a_\infty = 45,625$ days and $\Omega_{\rm x} := \cup\Omega_k$, $k = 1,\ldots,M$, where M being the number of states and union territories in India. We assume that there will be no inter- or intra-state and international movements, that is, $\mathbf{u} = 0$, no growth in level of infection, that is, $G_{\ell_v} = 0$, the growth age is negligible and no source. Nevertheless, population distribution in all three internal ordinates ℓ_v, ℓ_d and ℓ_a are described and tracked. Hence, the PBE in model (1) becomes

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial \ell_d} + CI = 0 \quad \text{in} \quad (0,\mathrm{T}_\infty] \times \Omega_{\rm x} \times \Omega_\ell. \tag{10}$$

Operator splitting finite element scheme

A finite element scheme[15] based on operator splitting[12,13,16] is used to solve the high-dimensional PBE model (10). Applying, operator splitting to (10), we get

Step 1. (x-direction) For given $I(t^a,x,\ell)$ with $I(t^a = 0,x,\ell) = I_0$, find $\tilde{I}(t^b,x,\ell)$ in (t^a,t^b) for all $\ell \in \Omega_\ell$ such that

$$\frac{\partial \tilde{I}}{\partial t} + CI = 0, \quad \tilde{I}(t^a,x,\ell) = I(t^a,x,\ell) \quad \text{in} \quad \Omega_{\rm x}\,, \tag{11}$$

Step 2. (ℓ_d-direction) For given $\tilde{I}(t^b,x,\ell)$, find $I(t^b,x,\ell)$ in (t^a,t^b) for all $x \in \Omega_{\rm x}$ $\ell_v \in L_v$ and $\ell_a \in L_a$ such that

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial \ell_d} = 0, \quad I(t^a,x,\ell) = \tilde{I}(t^b,x,\ell) \quad \text{in} \quad \Omega_\ell\,; \qquad \mathrm{I(t,\ell_v,0,\ell_a) = B_{nuc}} \quad \text{in} \quad \mathrm{L_a}, \tag{12}$$

In the x-direction, the evaluation equation (11) has to be solved for every $\ell \in \Omega_\ell$ by considering ℓ as a parameter. Similarly, the system (12) has to be solved in ℓ_d-direction for every $x \in \Omega_{\rm x}$ $\ell_v \in L_v$ and $\ell_a \in L_a$ by considering these variables as parameters. The backward Euler and discontinues Galerkin with upwind methods are used for the temporal and the spatial discretisation, respectively. The implementation of the splitting algorithm in the finite element context has been presented in these papers[12,13,16].

Parameters for Covid-19 predictions

The nucleation model B_{nuc} defined in (4) is considered with $R_0 = 3.35$, $f_1 = 1$, $f_2 = 1$, $f_4 = 1$ and

$$f_3(t,S_D) = 1. - 1./(1. + \exp(-(S_D(t) - 0.5)/0.1)), \quad S_D(t) = \begin{cases} 0.7 + 0.001333\,t & 0 \le t < 15 \\ 0.72 + 0.004285(t-15) & 15 \le t < 36 \\ 0.81 - 0.004(t-36) & 36 \le t < 51 \\ 0.75 + 0.0012(t-51.) & 51 \le t < 72 \\ 0.3 + ds(t-72) & \text{else} \end{cases}$$

Moreover, the values given in Table 1 are used for ds and f_1 to perform scenario analysis.

Furthermore, the quarantine function

$$\gamma_Q(\ell_d) = \begin{cases} 1 & 0 \le t < 1 \\ 0.9(1./(1. + \exp(-(l_v - 0.4)/0.1))) * (1./(1. + \exp(-(l_d - 5.1)/2))) & \text{else,} \end{cases}$$

Scenarios	Current Trend	Better Trend	Periodic lockdowns
ds	0.000333	0.0005	0.000333 & $f_1 = 0.01$ on lockdown days

Table 1. Parameter values used in scenario analysis.

is used in all scenarios. Finally, the recovery and death rate functions are fitted as

$$C_{ID}(t,\ell_v) = \begin{cases} 0 & \ell_v < 0.64 \\ 0.0475 - 0.000357\,t & \ell_v \geq 0.64 \text{ and } t < 21 \\ 0.0475 + 0.000208(t-21) & \ell_v \geq 0.64 \text{ and } 21 \leq t < 36 \\ 0.0475 & \text{else,} \end{cases}, \quad C_R(t) = \begin{cases} 0.01 + t * 0.00058 & t < 65 \\ 0.0475 & \text{else.} \end{cases}$$

Remark: A factor $N_S(t)/N(t)$ needs to be introduced in B_{nuc} when herd immunity develops.

Initial infection number density

For India, the data to estimate $N_D(\mathbf{x})$ is downloaded from a publicly sourced database[17]. To distribute the number density among the internal ordinates, we employ distribution fits as given in equation (9). Data set of age downloaded from Statista[18] is used to fit a_1, b_1, c_1 of equation (9).