



# City-Scale Agent-Based Simulators for the Study of Non-pharmaceutical Interventions in the Context of the COVID-19 Epidemic

IISc-TIFR COVID-19 City-Scale Simulation Team

Shubhada Agrawal<sup>2</sup>, Siddharth Bhandari<sup>2</sup>, Anirban Bhattacharjee<sup>2</sup>, Anand Deo<sup>2</sup>, Narendra M. Dixit<sup>1</sup>, Prahladh Harsha<sup>2</sup>, Sandeep Juneja<sup>2</sup>, Poonam Kesarwan<sup>2</sup>, Aditya Krishna Swamy<sup>1</sup>, Preetam Patil<sup>1</sup>, Nihesh Rathod<sup>1</sup>, Ramprasad Saphtharish<sup>2</sup>, Sharad Shriram<sup>1</sup>, Piyush Srivastava<sup>2</sup>, Rajesh Sundaresan<sup>1\*</sup>, Nidhin Koshy Vaidhiyan<sup>1</sup> and Sarath Yasodharan<sup>1</sup>

**Abstract** | We highlight the usefulness of city-scale agent-based simulators in studying various non-pharmaceutical interventions to manage an evolving pandemic. We ground our studies in the context of the COVID-19 pandemic and demonstrate the power of the simulator via several exploratory case studies in two metropolises, Bengaluru and Mumbai. Such tools may in time become a common-place item in the tool kit of the administrative authorities of large cities.

## 1 Introduction

COVID-19 is an ongoing pandemic that began in December 2019. The first case in India was reported on 30 January 2020. The number of cases and fatalities have been on the rise since then. As on 21 September 2020, there have been 54,87,580 confirmed cases (of which 43,96,399 have recovered) and 87,882 fatalities<sup>1</sup>; see Fig. 1 for a timeline of COVID-19 cases, recoveries, and fatalities in India. While medicines/vaccines for treating the disease remained under development at the time of writing this paper, many countries implemented non-pharmaceutical interventions such as testing, tracing, tracking and isolation, and broader approaches such as quarantining of suspected cases, containment zones, social distancing, lockdown, etc., to control the spread of the disease. For instance, the Government of India imposed a nation-wide lockdown from 25 March 2020 to 14 April 2020, and subsequently extended it until 31 May 2020 to break the chain of transmission and also to mobilise resources (increase healthcare facilities and streamline procedures). To evaluate various such interventions

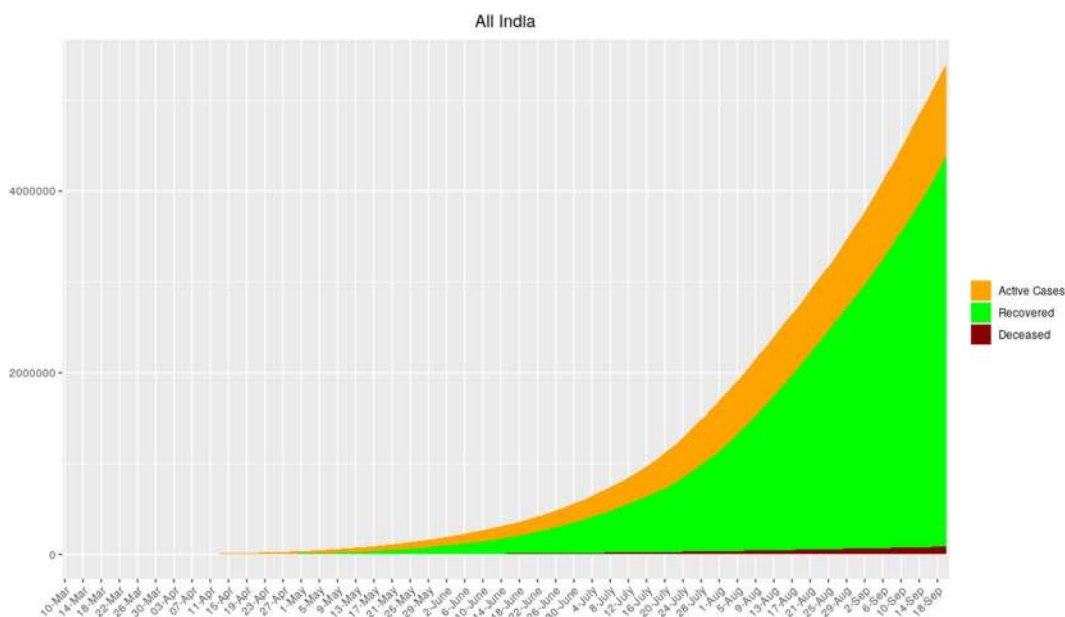
and decide which route to take to manage the pandemic, epidemiologists resort to models that predict the total number of cases and fatalities in both the immediate and the distant futures. The models used should have enough features to enable the evaluation of the impact of various kinds of non-pharmaceutical interventions.

Broadly, three kinds of models have been used to study this epidemic. The first set of models takes a curve-fitting approach. They rely on simple parametric function classes. The parameters of the model are fit via regression to match observed trends. The second set of models addresses the physical dynamics of the spread at a macroscopic level. These are mean-field ordinary differential equations (ODEs)-based compartmental models (e.g., Susceptible-Exposed-Infected-Recovered (SEIR) model and its extensions) based on the classical work of Kermack and McKendrick<sup>2</sup>. Here, the population is divided into various compartments such as susceptible, exposed, infected, recovered, etc., based on the characteristics of the epidemic. One then solves a system of ODEs that captures the evolution of the epidemic at a

Source code available at: <https://github.com/cni-iisc/epidemic-simulator/releases/tag/3.0>

<sup>1</sup> Indian Institute of Science, Bangalore, India.

<sup>2</sup> TIFR, Mumbai, India. [rajeshs@iisc.ac.in](mailto:rajeshs@iisc.ac.in)



**Figure 1:** Timeline of COVID-19 cases, recoveries, and fatalities in India taken from<sup>11</sup>. See<sup>1</sup> and<sup>12</sup> for detailed information on how COVID-19 progressed in India.

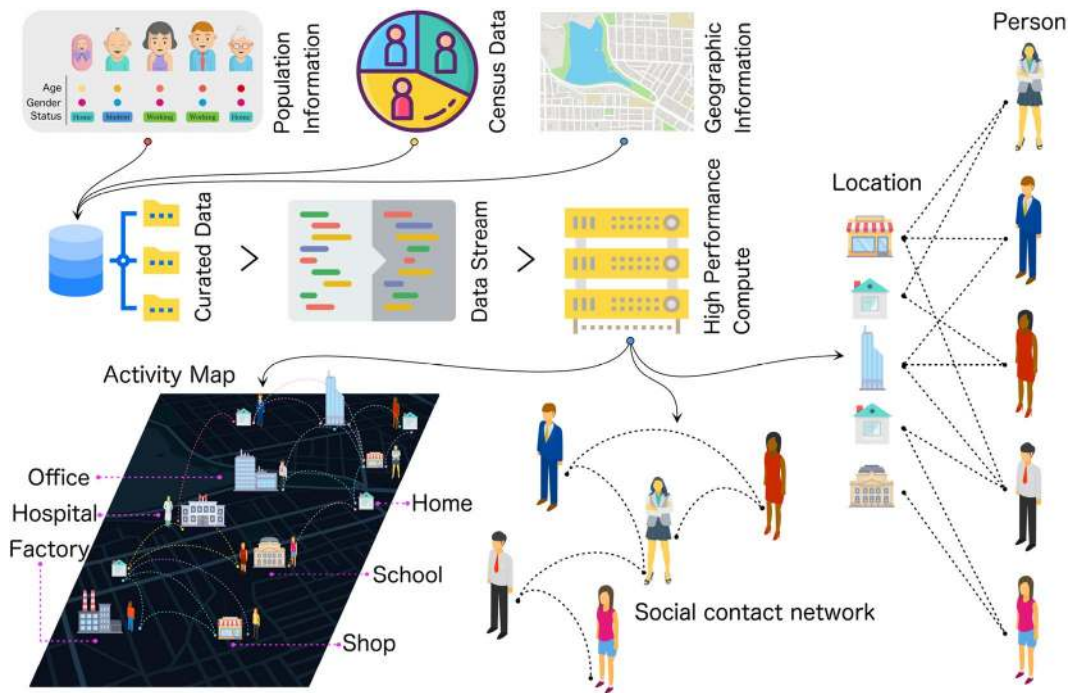
macroscopic scale.<sup>1</sup> Localised versions of these are spatio-temporal mean-field models that lead to partial differential equations.<sup>2</sup> The third set of models, and the focus of this work, are agent-based models.<sup>3</sup> A very detailed model of the society under consideration, with as many agents as the population, is constructed using census and other data. The agents interact in various interaction spaces such as households, schools, workplaces, marketplaces, transport spaces, etc. See Fig. 2 for a schematic representation of an agent-based model with the aforementioned interaction spaces. These interaction spaces are the primary contexts for the spread of infection. A susceptible individual can potentially get infected from an interaction in one of these spaces upon contact with an infected individual. Once an individual is exposed to the virus, this person goes through various stages of the disease, may infect others, and, eventually, either recovers or dies. Other models work at an intermediate level by modelling the social network of interactions, e.g.,<sup>10</sup>, but we shall focus more on agent-based models.

<sup>1</sup> See<sup>3</sup> for a state-level epidemiological model for India and<sup>4</sup> for a combination of the two approaches.

<sup>2</sup> For a paper in the Indian context, see<sup>5</sup>.

<sup>3</sup> There are other agent-based simulators that have informed policy decisions. See<sup>6</sup> for UK and USA-related studies specific to COVID-19, see<sup>7</sup> for a COVID-19 study on Sweden, see<sup>8</sup> and references therein for many agent-based models and their comparisons, and see<sup>9</sup> for a taxonomy of agent-based models.

There are several advantages of using agent-based models. First, since modelling is performed at a microscopic level unlike the macroscopic level in compartmental models, agent-based models are well suited to capture heterogeneity at various levels. For instance, the age-dependent progression of COVID-19 in individuals (severity, the need for hospital care, intensive care, etc.) can be incorporated in agent-based models. Second, individual behavioural changes, known to be important in certain diseases such as AIDS, can be easily modelled. Third, agent-based models are well suited to study the impact of various non-pharmaceutical interventions, such as “lockdown for a certain number of days”, “offices operating using the (so-called) odd-even strategy”, “social distancing of the elderly”, “voluntary home quarantine”, “closure of schools and colleges”, etc. Explicit modelling of these contexts of infection spread also enables studies of control measures targeting the interaction spaces. Fourth, there is an important difference between the *actual infected number* in the population, which is what the differential equations-based models predict, and the *reported cases*. The latter is invariably based on those that come to hospitals/clinics seeking health care, or those that are identified due to random testing, followed by contact tracing of such index cases. As a consequence, reported cases provide a *biased* estimate of the actual infected number in the population.



**Figure 2:** Schematic representation of an agent-based model.

Agent-based simulators have the capability to track such biased estimates of prevalence.

In this work, we describe our city-scale agent-based simulator to study the epidemic spread in two Indian cities and demonstrate how digital computational capabilities can help us to assess the impact of various interventions and manage a pandemic.

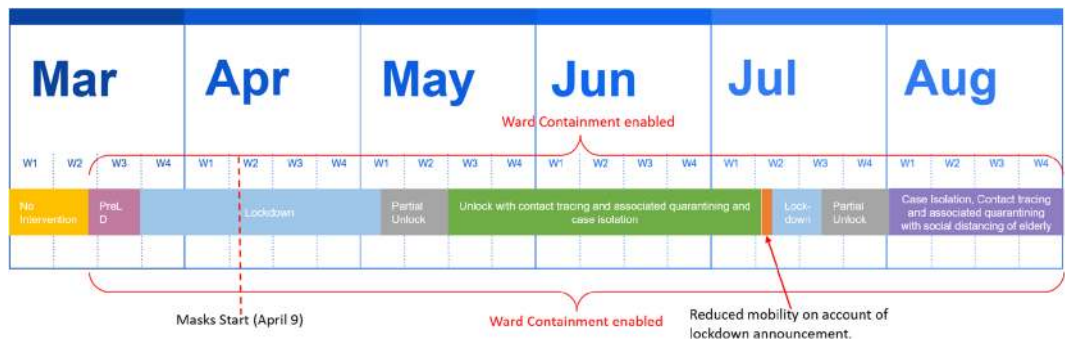
We now provide sample outcomes for Bengaluru and Mumbai for COVID-19 under various interventions. These outcomes have been generated using our city-scale agent-based simulator. Bengaluru and Mumbai have estimated populations of 1.23 crore (12.3 million) and 1.24 crore (12.4 million) people, respectively,<sup>4</sup> and our simulator has instantiated those many agents. The Bengaluru population is spread over 198 administrative units called wards. Similarly, the Mumbai population is spread over similar administrative units or wards, but there are 24 such wards in Mumbai. Since these are larger wards compared

to Bengaluru's wards, there is significant variation of population density within each ward in Mumbai. To model higher spread in densely populated areas, each of the 24 wards is modelled to have subareas with denser population.

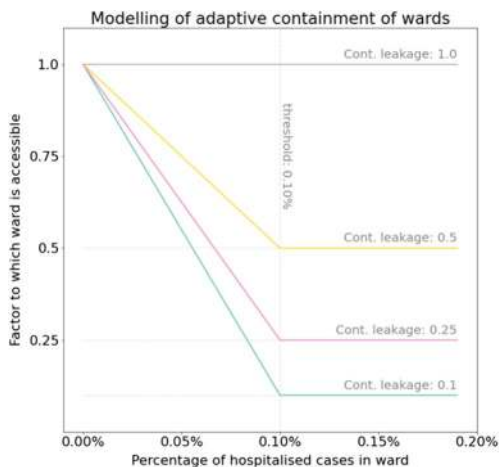
1. *Bengaluru* For Bengaluru, we consider the following scenario. The Government of India implemented a 40-day lockdown starting 25 March 2020. Bengaluru and Karnataka had already closed some interaction spaces in the form of a pre-lockdown. After the 40-day lockdown, there was a phased opening of various activities and offices. More details are as follows. These can also be read from the bottom timeline of Fig. 3. We then provide a simulation outcome for these interventions and compare them with the actual situation on the ground. The details:

- Pre-lockdown from 14 to 24 March 2020, the first shaded region in Figs. 5, 6.
- 40 days of lockdown from 25 March to 03 May 2020, the second and the third shaded regions in the figures.
- 14 days of phased opening from 04 to 17 May 2020 involving voluntary home quarantine, social distancing of the elderly, closure of schools and colleges, 50% occupancy at workplaces, and case isolation. This is the fourth shaded region in the plots.

<sup>4</sup> The 2011 census figure for Bengaluru is 0.85 crore (8.5 million) and that for Mumbai is 1.24 crore (Mumbai city only, not the Mumbai Urban Area whose 2011 census estimate is 1.84 crore (18.4 million)). Bengaluru's 2020 population is estimated to be 1.23 crore (12.3 million). Reliable data are not available for Mumbai city's 2020 population. We have used 2020 estimated population for Bengaluru and 2011 census estimate for Mumbai.



**Figure 3:** A timeline of Bengaluru interventions.



**Figure 4:** Wards are contained in a 'soft' way. The mobility is gradually decreased based on the signal of number of hospitalised cases in the ward. When 1 in 1000 in the ward is hospitalised, a local lockdown comes into effect.

- From 18 May 2020 onwards, continued contact tracing (following the Indian Council of Medical Research (ICMR) guidelines as much as possible) and associated quarantining and case isolations, but otherwise an unlocked Bengaluru. *Soft ward-containment* continues to be in force. By soft ward-containment, see Fig. 4, we mean linearly varying mobility control that turns an open ward into a locked ward when the number of hospitalised cases becomes 0.1% of the ward's population; in the latter locked scenario, only 25% mobility is allowed for essential services.
- Past studies<sup>13–16</sup> have indicated that masks have been effective in reducing the spread of influenza. Anecdotal evidence seems to suggest that masks are effective for COVID-19. The Ministry of Home Affairs (MHA) order of 15 April 2020<sup>17</sup>, Annexure 1 made the wearing of masks in public places compulsory.

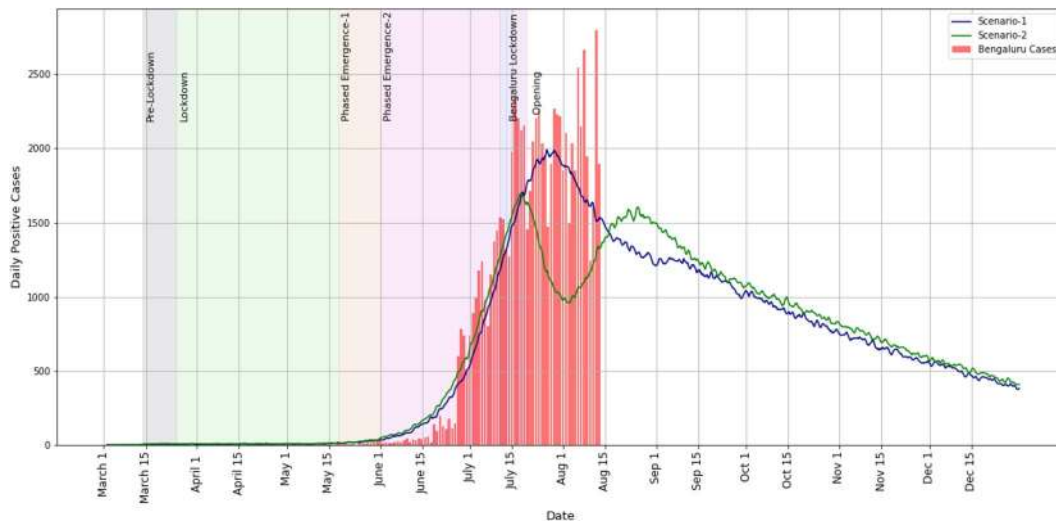
This was re-emphasised in the MHA order of 30 May 2020<sup>18</sup>. We assume that masks are mandatory from 09 April 2020 onwards.

- It is often the case that when there are several restrictions in place, only a fraction of the population complies with these restrictions. Getting the entire population to comply is often a big challenge and requires significant and persistent messaging (including communication, rewards, and punitive measures). We assume a *compliance factor* of 0.7 up to 04 May 2020, which means that 70% of the population adheres to the government guidelines like social distancing, wearing masks in public places, etc., and 0.6 thereafter. The reduction could be attributed to behavioural changes due to lockdown fatigue.
- A brief lockdown during 14–21 July 2020 was implemented in Bengaluru. We compare two scenarios, one without this lockdown (we call this Scenario-1) and one with this lockdown (we call this Scenario-2, see Fig. 3).

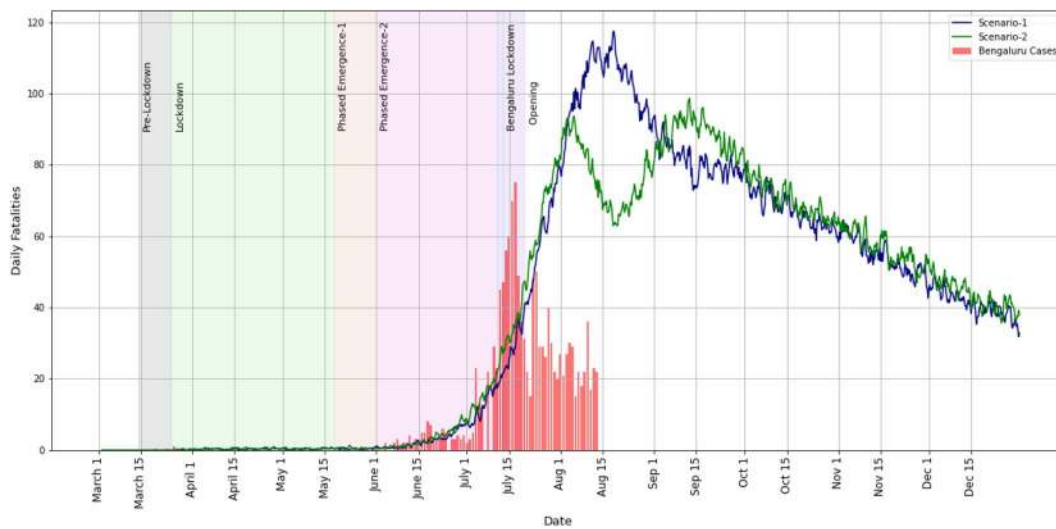
As one can anticipate, simulation of the above scenarios requires a significant level of sophistication in the modelling and implementation. We describe how we do these in the coming sections, but now focus only on the outcomes.

Figure 5 estimates the daily positive cases and Fig. 6 estimates the daily fatalities directly due to COVID-19 in Bengaluru. In both these plots, we compare our estimates with the situation on the ground. The plots are the means of five runs each on two versions of synthetic Bengaluru. The jaggedness is due to the stochasticity associated with the limited number of runs. For greater clarity, we have not included the standard error plots.

Figures 5 and 6 provide the estimates with and without this 1-week lockdown. The trend for the reported cases is roughly captured, but fatalities are over-predicted. This is surprising,



**Figure 5:** Bengaluru daily positive cases estimation. The red bars are the reported cases. The five shaded regions between 14 March and 01 June represent the durations of the various lockdown phases. The shaded region around 15 July represents a short 1-week lockdown. For cumulative case plots, see Fig. 11.



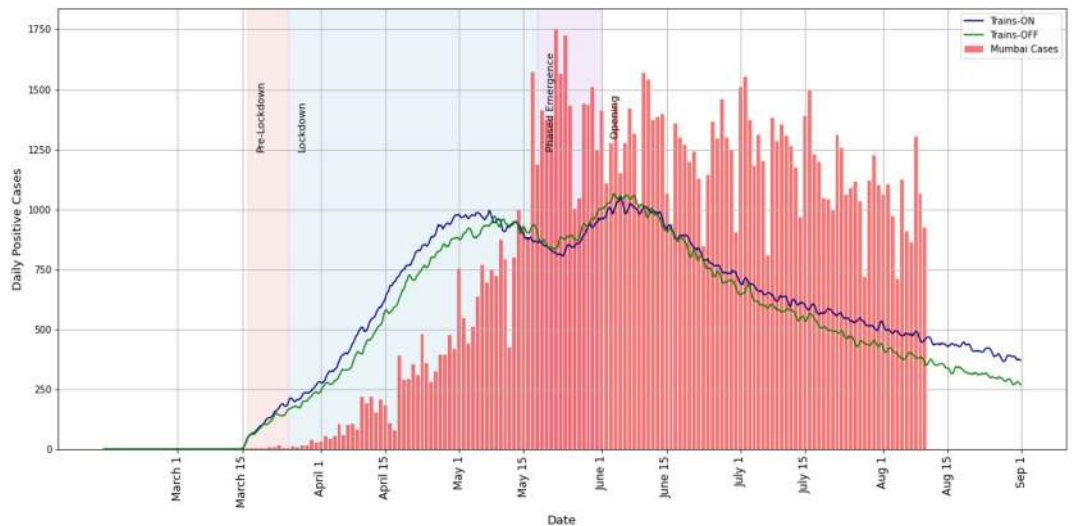
**Figure 6:** Bengaluru daily fatalities' estimation. For cumulative fatalities, see Fig. 12 in a later case study.

since the reported cases continued to be high in the third week of July. For a more detailed study of these plots, we refer the reader to Sect. 3.1. At this stage, we only observe that the public-health benefit of the lockdown is clear from the pictures, reduced peak at the expense of a brief second wave. Armed with these predicted outcomes under the two scenarios, public-health officials can now weigh the benefits of the lockdown against its economic consequences.

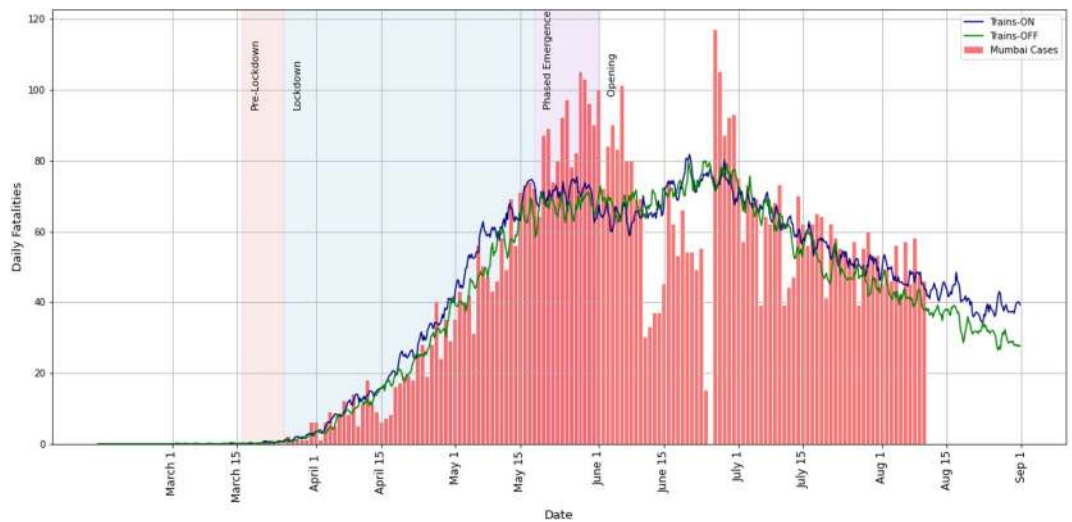
2. *Mumbai* For Mumbai, we simulate the following scenario.

- A pre-lockdown similar to Bengaluru, but during 16–24 March 2020 (first shaded region in Figs. 7, 8).
- 40 + 14 days of lockdown from 25 March to 17 May 2020, the second, third, and fourth shaded regions in the figures.
- Masks are mandatory from 09 April 2020.
- Workplaces open with a small strength of 5% during 18–31 May 2020, as per Government of Maharashtra directions. This is the fifth shaded region. During this period, social distancing of the elderly and school and college closures remain in force.





**Figure 7:** Mumbai daily positive cases' estimation. The five shaded regions between 16 March and 01 June represent the durations of the various lockdown phases in Mumbai. For cumulative case plots, see Fig. 21.



**Figure 8:** Mumbai daily fatalities' estimation vs. corrected fatality time series, as corrected by BMC on 18 June 2020. For cumulative fatalities, see Fig. 22 in a later case study.

- Workplace strengths increase to 20% in June, to 33% in July, and to 50% in August, with commensurate capacity increases in the local trains. Social distancing of the elderly and school and college closures remain in force. In addition voluntary home quarantine and case isolation come into play.
- Throughout the simulations, soft ward-containment is in force (with 25% mobility allowed in the lockdown scenario, see Fig. 4).
- It is often difficult to comply with social distancing directives in high population density areas, like in slums, with many common essential facilities. In Mumbai, we model compliance to be 0.4 in high-density areas and 0.6 in other areas.
- Throughout the simulations, contact tracing, associated quarantining, testing, and further tracing are enabled.
- We will compare the above scenario, with local trains enabled, and will contrast it with another hypothetical scenario having no local trains.

Figure 7 estimates the daily positive cases and Fig. 8 estimates the daily fatalities. In both these

plots, we compare our estimates with the situation on the ground. Again, the plots are the means of five runs each on two synthetic versions of Mumbai.

Except for a delay in the estimated cases curve, the trends for cases and fatalities are captured well. The delay in the estimated cases is perhaps due to delayed reporting which is not modelled in the simulator. From these figures, one can recognise the usefulness of the agent-based simulator in assessing the impact of opening of the local trains in Mumbai.

*Agent-based simulator—an important intervention-planning tool* Let us summarise the outcomes of the above two examples. We saw the public-health impact of imposing a short lockdown in Bengaluru. We also saw the impact of opening up trains vs. keeping the trains non-operational in Mumbai. Such comparisons that can inform better on-ground decision-making are enabled by a city-scale agent-based simulator. This capability arises mainly because of the enhanced modelling and control of the interaction spaces in the simulator. It is our hope that such tools become common place in a city administration's tool kit, and are used to the fullest extent before drastic interventions with wide-scale impact, e.g., lockdown, are imposed. With additional modelling of activity, mobility, and behaviour, and use of high-quality data on the migrant labour force in urban areas, we speculate that we could have anticipated certain behavioural outcomes seen in India after the lockdown announcement (e.g., migrant population movement).

## 2 Methodology: Agent-Based Modelling

Broadly, the steps involved in agent-based modelling are the following: build the simulator, calibrate it, validate it, and use it for estimating how the pandemic will evolve.

1. *Simulator* The simulator itself consists of four parts.

*Synthetic city* A synthetic city generator builds a synthetic city with individuals and various interaction spaces. Individuals are assigned to various interaction spaces such as households, schools/workplaces, communities, and transport spaces. In doing this, we capture the demographics of the city, the school-size distributions, the workplace size distributions, the commute distances, the neighbourhood and friends' interaction networks, the transport interaction spaces, etc. These fix the "social networks" on which individuals interact and transmit the virus.

*Disease progression* A disease progression model that involves the biology of the disease then indicates what is the incubation period, infective period, symptomatic period, severity of the symptoms, viral load, virus shedding, health care and in-hospital progression, etc.

*Interactions and evolution* The level of infectivity during the infective period, the duration of the infective period, and the social network interactions in the various interaction spaces determine how the disease evolves in the city. We start the simulator with a certain number of infected individuals. They then interact with susceptible individuals at various interaction spaces, who in turn interact with other susceptible individuals, and thus the epidemic progresses. The key parameters in the disease evolution are the transmission coefficients associated with each interaction space that model the chance of meetings and disease spread in that interaction space.

*Intervention model* Various kinds of intervention policies need to be defined and their impact on transmission coefficients should be modelled. See Table 1 for some examples. Many of these involve reduction in changes in contact rates as a consequence of the interventions. The values to set could be based on observed mobility patterns. For example, according to the COVID-19 Community Mobility Report for India in April<sup>19</sup> in Table 2, prepared by Google based on data from Google Account users who have "opted-in" to location history, there was significant reduction in mobility during the lockdown period compared to the baseline period of 03 January 2020–06 February 2020. This informs the nominal contact rate choices in the interventions' definitions in Table 1 and later in other Tables.

2. *Calibration* Once the simulator is ready, there are still unknown parameters that need to be identified. These include the contact rates at various interaction spaces, the number of infections to seed, the time at which these infections should be seeded, the compliance parameters, etc. The purpose of the calibration step is to identify these parameters to capture the city specific trends and contact rates. We do this by choosing the initial number to seed, the time at which these are seeded, and the contact rates, so that the initial trend of the disease is matched. Once calibrated, we can run our simulator for a certain number of days and understand how the epidemic spreads.

3. *Validation* We next have to validate our simulator, so that we can understand the predictive power of the simulator. For this, we look for phenomena in the real data that have not been

**Table 1:** Intervention modelling.

Label	Policy	Description
CI	Case isolation at home	Symptomatic individuals stay at home for 7 days, non-household contacts reduced by 90% during this period, household contacts reduce by 25%
HQ	Voluntary home quarantine	Once a symptomatic individual has been identified, all members of the household remain at home for 14 days. Non-household contacts reduced by 90% during this period, household contacts reduce by 25%
SDO	Social distancing of those aged 65 and over	Non-household contacts reduce by 75%
LD	Lockdown	Closure of schools and colleges. Only essential workplaces active. For a compliant household, household contact rate doubles, community contact rate reduces by 75%, workspace contact rate reduces by 75%. For a non-compliant household, household contact rate increases by 25%, workspace contact rate reduces by 75%, and no change to community contact rate

**Table 2:** Mobility report generated on 11 April 2020, see<sup>19</sup>.

Place	Reduction (%)
Retail and recreation	– 80
Grocery and pharmacy	– 55
Parks and public plazas	– 52
Public transit stations	– 69
Workplaces	– 64
Residential	+ 30

explicitly modelled and we check if the simulator is able to capture these phenomena. For specific details, see Sect. 4.

#### 4. Use of the simulator in an evolving pandemic

It is often the case that in evolving pandemics, predictions do not match reality as time unfolds. Models are often gross over-simplifications of the underlying complex reality and assumptions are often wrong or may need updating as the pandemic evolves. The purpose of models in an evolving pandemic is not merely to predict numbers, in which task they will likely fail, but more to enable principled decision-making on intervention strategies. They enable a comparative study of the public-health outcomes of one strategy versus another. Armed with these comparisons, public-health officials can make more informed decisions. Needless to say, these are often more complex and involve several aspects beyond just public health, e.g. economy, psychology, education, political climate, to name a few.<sup>5</sup>

<sup>5</sup> For a proposal on how to simulate economic and public-health aspects together, see<sup>20</sup>.

### 3 Design of Interventions via Case Studies

One of the powerful features of the agent-based simulator is its ability to explicitly control various interaction spaces and study the outcomes. We demonstrate this feature via the case studies for Bengaluru and Mumbai listed in Table 3.

#### 3.1 Case Study A: No Intervention (But Only Contact Tracing-Based Isolation) Versus Lockdown Versus the Current Scenario in Bengaluru

We compare the following three scenarios in Bengaluru:

- No intervention other than contact tracing, testing, and associated case isolation.
- Indefinite lockdown starting from 14 March 2020 onwards. This naturally will have enormous economic and societal cost, but we focus only on the direct COVID-19 public-health outcomes.
- Scenario-2 in Table 4: soft ward-containment, case isolation with testing and contact tracing, and a 1-week lockdown during 14–21 July 2020.

We assume a compliance of 70% until 03 May 2020 (i.e., during the initial Karnataka-wide lockdown followed by the nation-wide lockdown) and a compliance of 60% starting 04 May 2020, for all these scenarios. That is, 70% (resp. 60%) of the population comply with the restrictions in place until 03 May 2020 (resp. starting 04 May 2020). Under these scenarios, we plot the following: daily cases (Fig. 9), daily fatalities (Fig. 10), cumulative cases (Fig. 11), cumulative fatalities (Fig. 12), and estimated hospital beds and critical care beds (Fig. 13). We make the following observations.



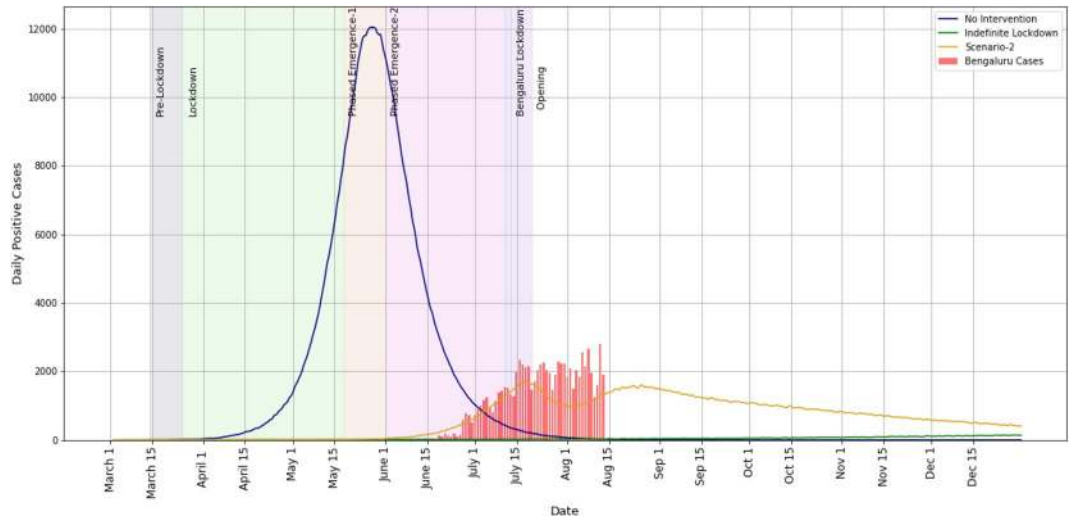
**Table 3:** Case studies for Bengaluru and Mumbai.

Case Study	Title	Section
Case Study A	No intervention (but only contact tracing-based isolation) versus lockdown versus the current scenario in Bengaluru	Section 3.1
Case Study B	Impact of opening offices at 50% capacity with higher compliance versus lockdown at lower compliance, a Bengaluru study	Section 3.2
Case Study C	Impact of opening trains versus not opening trains in Mumbai	Section 3.3
Case Study D	Soft ward-containment versus neighbourhood containment in Bengaluru	Section 3.4
Case Study E	Soft ward-containment at various levels in Mumbai	Section 3.5
Case Study F	Schools/colleges open from 01 September 2020 in Bengaluru	Section 3.6

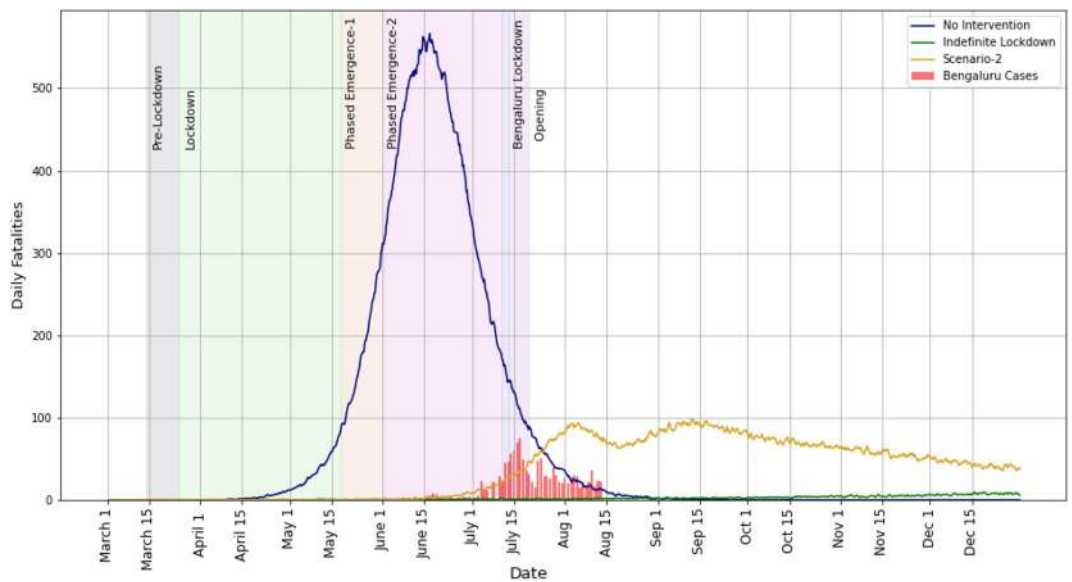
**Table 4:** Simulated Bengaluru interventions.

Period	Scenario-1	Scenario-2	Compliance (%)
01–13 March 2020		No intervention	NA
14–24 March 2020		Pre-lockdown	70
25 March–03 May 2020		40 days of National lockdown	70
09 April 2020–onwards		Masks ON	70
04–17 May 2020	Phased opening. Voluntary home quarantine, social distancing of elderly, case isolation, schools and colleges closed, 50% occupancy at workplaces		60
18 May–11 July 2020	Unlocked Bengaluru with only ICMR-guideline contact tracing and associated quarantining and case isolations, social distancing of elderly, case isolation, schools and colleges closed		60
12–13 July 2020	Same as above	Pre-lockdown—50% mobility	60
14 July–21 July 2020	Same as above	Bengaluru lockdown	60
22 July–31 July 2020	Same as above	Case isolation, social distancing of elderly, school closure, workplaces at 50%	60
01 August 2020–onwards	Same as above	Case isolation, social distancing of elderly, school closure, workplaces at 100%	60
Throughout	Soft ward-containment enabled		60

- As one would expect, the least number of cases, fatalities, and hospital beds' requirements correspond to the "indefinite lockdown" scenario. However, this scenario has serious impact on the economy, livelihoods, etc.
- In terms of the daily number of cases, the no-intervention scenario had a peak around 01 June 2020 (with roughly 15,000 cases), whereas the present scenario in Bengaluru (i.e., Scenario-2 in Table 4) had a much lower peak around 15 July 2020 (with around 2000 cases), followed by another peak around end of August. Similar trends can be seen in the fatalities estimates as well as the hospital bed estimates. Our health care system would have struggled with the no-intervention scenario, and the present scenario in Bengaluru helped to mitigate and delay the peak of the epidemic.
- The second predicted peak in Scenario-2 in Table 4 is due to the 1-week lockdown during 14–21 July 2020.
- Towards the end of July, we over-predict the number of daily fatalities and underpredict the number of daily cases. This could be because of two reasons:
  1. The number of tests has increased significantly during mid-July due to which there is a likely surge in the number of asymptomatic cases. As a consequence, a reduction in the number of daily



**Figure 9:** Case study A, Sect. 3.1: Bengaluru daily cases' estimation. For a magnified view of the lower part of the plot, see Fig. 5. The no-intervention situation would have overwhelmed the healthcare system many times over.

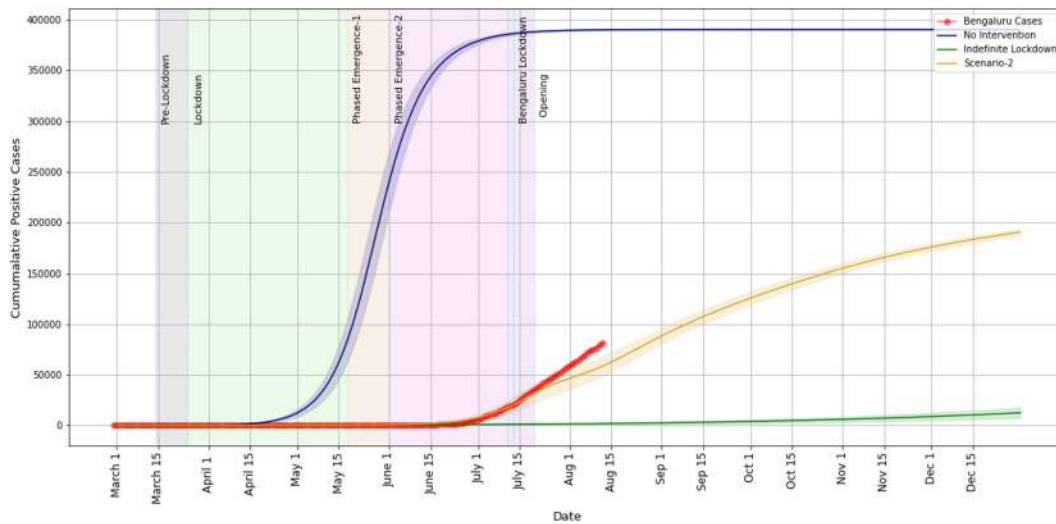


**Figure 10:** Case study A, Sect. 3.1: Bengaluru daily fatalities' estimation. For a magnified view of the lower part of the part, see Fig. 6.

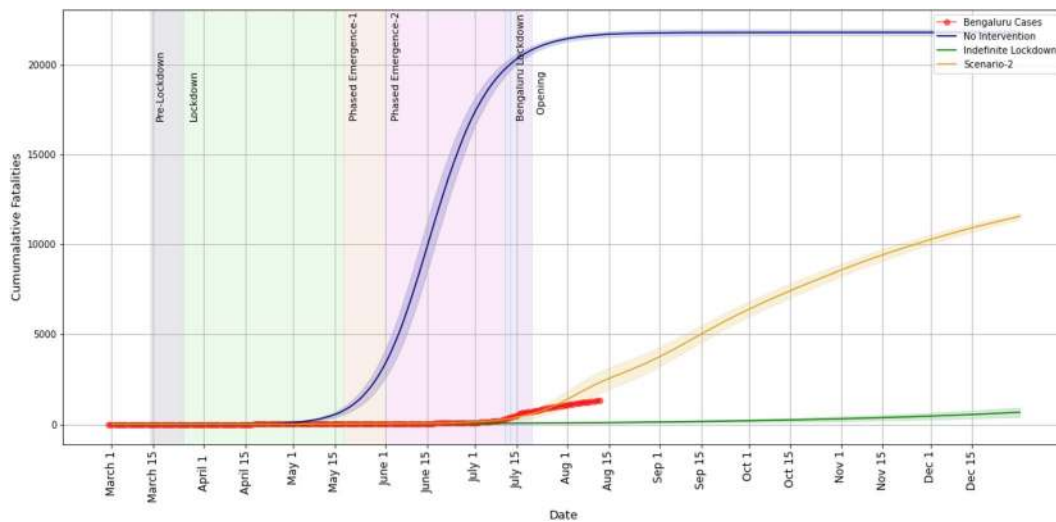
cases due to the 1-week lockdown during 14–21 July is not observed in the reported number of daily cases; such a reduction is visible in our estimates, because the testing regime is assumed constant through the period in our simulator.

2. There is a delay in reporting the fatalities. As the reported number of daily cases follow an exponential trend during early–mid July, one would expect a

similar trend in the reported daily fatalities during end-July, as shown in our prediction of the daily fatalities under Scenario-2. However, we see a reduction in the reported number of daily fatalities during after 15 July 2020. This could be due to a possible delay in reporting the daily fatalities, or an effective use of the rapid point-of-care antigen test kits, or a combination of both. Testing of these hypotheses requires further investigation.



**Figure 11:** Case study A, Sect. 3.1: Bengaluru cumulative cases' estimation.



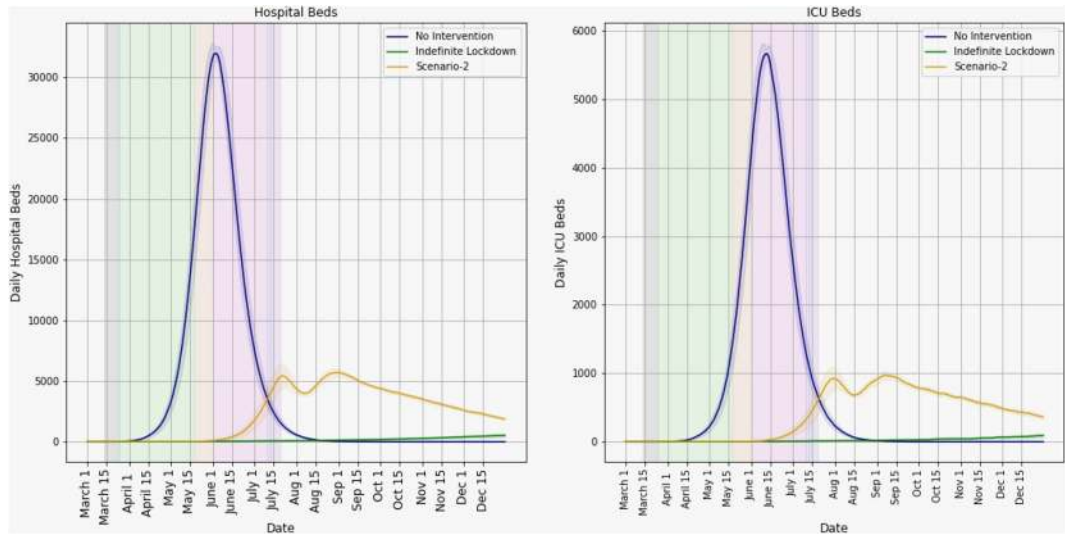
**Figure 12:** Case study A, Sect. 3.1: Bengaluru cumulative fatalities' estimation.

### 3.2 Case Study B: Impact of Opening Offices at 50% Capacity with Higher Compliance Versus Lockdown at Lower Compliance

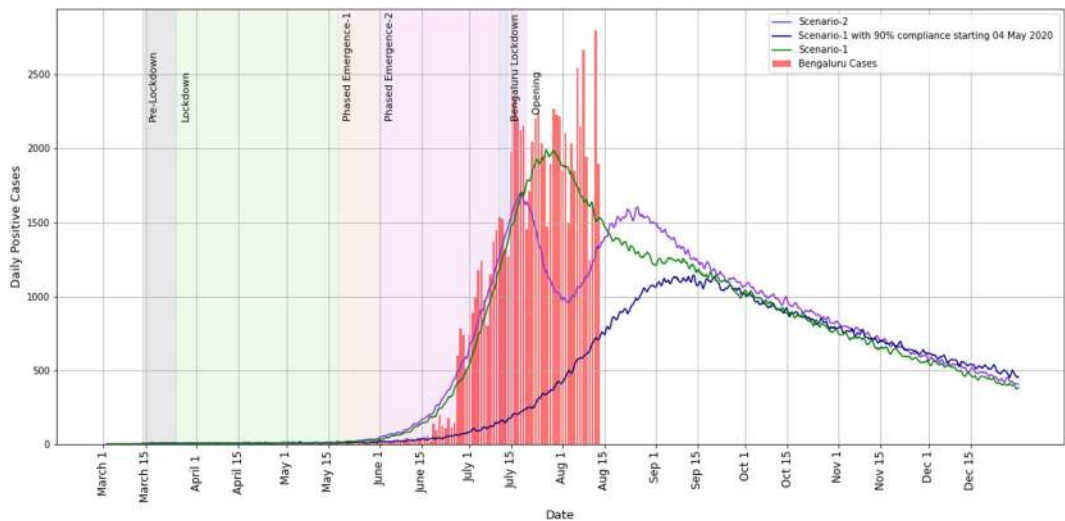
The degree of compliance among the population to public-health directions/guidelines is an important factor that affects the epidemic. To understand the importance of compliance, we compare the following scenarios for Bengaluru: the present Bengaluru (i.e., Scenario-2 in Table 4), an unlocked Bengaluru (i.e., Scenario-1 in Table 4), and an unlocked Bengaluru with a higher compliance of 90% starting 04 May 2020 (i.e., Scenario-1 in Table 4 with 70% compliance during 14 March 2020–03 May 2020 and 90%

compliance starting 04 May 2020). As before, we plot the following: daily cases (Fig. 14), daily fatalities (Fig. 15), cumulative cases (Fig. 16), cumulative fatalities (Fig. 17), and estimated hospital beds and critical care beds (Fig. 18). We make two important observations:

- In terms of the number of cases and fatalities, the present Bengaluru (i.e., Scenario-2 in Table 4) with 60% compliance starting 04 May 2020 is worse than an unlocked Bengaluru with 90% compliance starting 04 May 2020 (with both scenarios having 70% compliance until 03 May 2020). While the qualitative outcome is not surprising, the quanti-



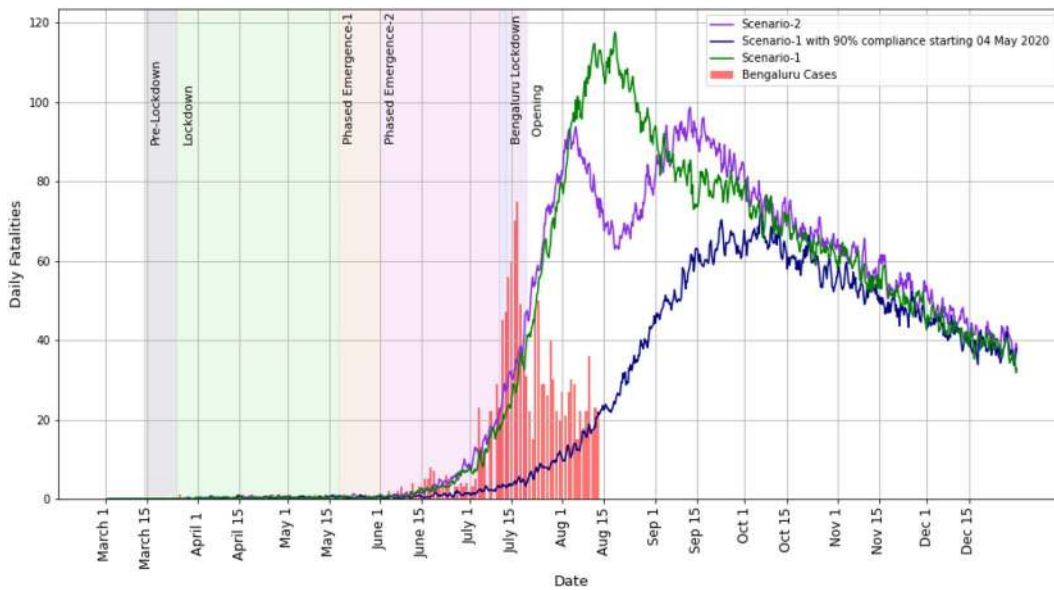
**Figure 13:** Case study A, Sect. 3.1: Bengaluru hospital beds' estimation. 'Hospital Beds' refers to the number of beds occupied for regular care including possibly oxygen support. 'ICU Beds' refers to those that need intensive care or ventilation. The no-intervention scenario would have overwhelmed Bengaluru's healthcare system.



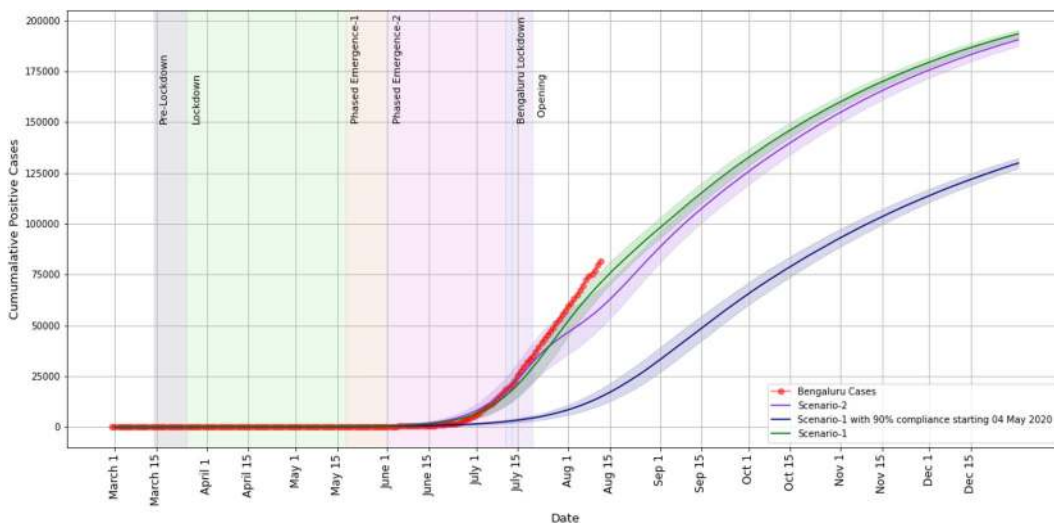
**Figure 14:** Case study B, Sect. 3.2: Bengaluru daily cases' estimation.

tative estimates suggest just how important compliance is in curbing the spread of the disease. Armed with such comparisons, city administrations could suitably allocate resources for communication, awareness, and other such campaigns to educate the general populace on the public-health impact of their actions, to induce more pro-social behaviour, and to ensure greater compliance. This was the approach taken by Sweden, a country with a population of about 1 crore (10 million).

- Comparing Scenario-1 and Scenario-2, we see that the effect of the 1-week lockdown during 14–21 July is very minimal in the long term as far as the cumulative number of cases and fatalities are concerned. However, there is a significant difference in the cumulative number of cases and fatalities between Scenario-2 and Scenario-1 with a higher compliance of 90% starting 04 May 2020. This suggests that, given that vaccines for COVID-19 are not yet available, short-term lockdowns' benefit is restricted to



**Figure 15:** Case study B, Sect. 3.2: Bengaluru daily fatalities' estimation.



**Figure 16:** Case study B, Sect. 3.2: Bengaluru cumulative cases' estimation.

mobilising resources and preparing the health-care system in the short term. On the other hand, higher compliance has a greater impact in reducing cases and fatalities.

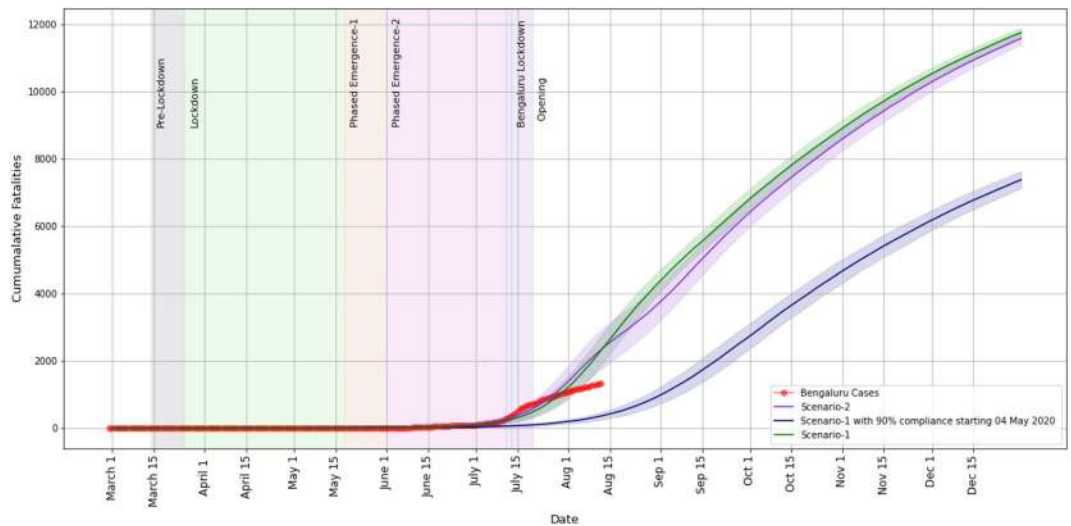
### 3.3 Case Study C: Impact of Opening Trains Versus Not Opening Trains in Mumbai

We now study the impact of opening suburban trains in Mumbai. Table 5 shows the timeline of various restrictions implemented in Mumbai

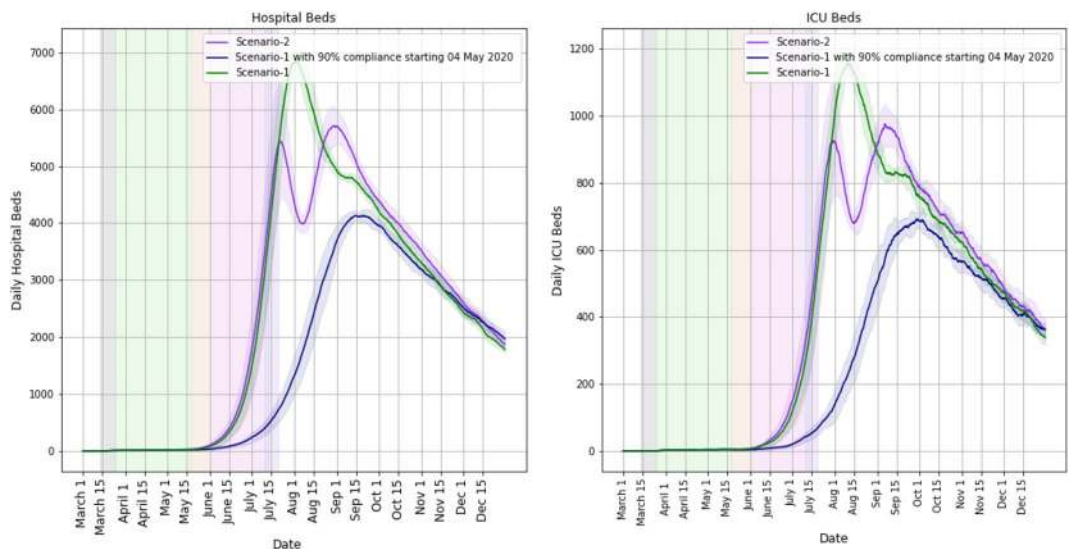
starting from 01 March 2020. Suburban trains were not under operation in Mumbai during 15 March–31 May 2020. As suburban trains are a key mode of daily commute in Mumbai, we compare the situation in Table 5 under two scenarios:

- **Trains-ON:** Suburban trains are operational starting 01 June 2020 in a phased manner, similar to the opening of workplaces in a phased manner as indicated in Table 5,
- **Trains-OFF:** Suburban trains are not operational throughout.





**Figure 17:** Case study B, Sect. 3.2: Bengaluru cumulative fatalities' estimation.



**Figure 18:** Case study B, Sect. 3.2: Bengaluru hospital beds' estimation.

As indicated in Table 5, we assume a compliance factor of 60% in non-slums and 40% in slums. We plot our results in Figs. 19, 20, 21, 22, and 23.

- From the plots, we see that the phased opening of suburban trains starting 01 June 2020 gives a marginal increase in the number of cases, fatalities, and hospital beds compared to the Trains-OFF scenario. This suggests that trains can be operated with enforcement of strict physical distancing (by operating at

reduced passenger loads)<sup>6</sup> and compulsory wearing of face masks.

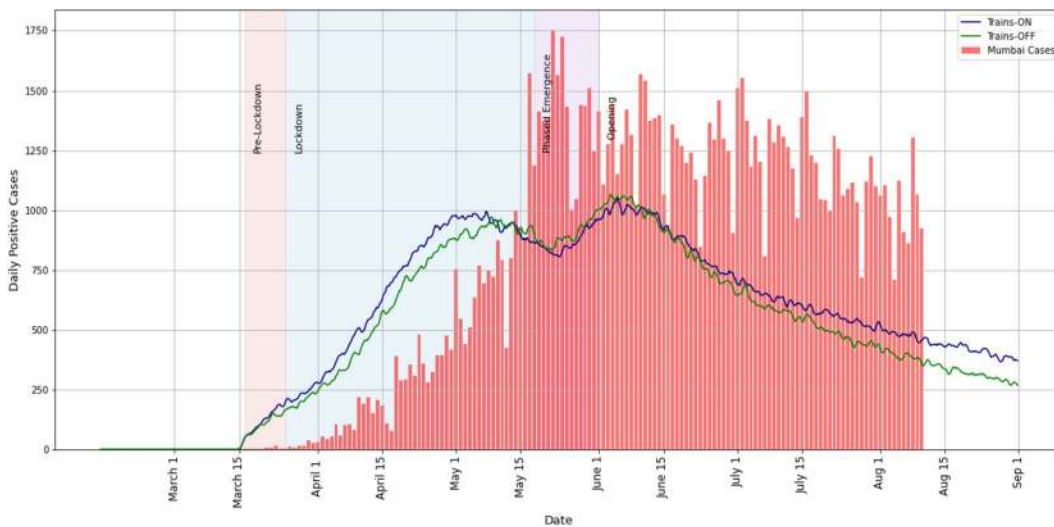
- Although we match the daily fatalities<sup>7</sup> curve very well, we over-predict the daily number of cases. We believe that this is due to the limi-

<sup>6</sup> Physical distancing under normal passenger loads in Mumbai locals is not possible given the large number of commuters per train.

<sup>7</sup> We use a corrected version of the reported number of daily fatalities from Brihanmumbai Municipal Corporation (BMC). This correction was done by BMC. The initial reported daily fatalities' curve from BMC had a very large peak at 16 June 2020. The corrected data adjust the daily fatalities curve until 15 June 2020, so that the peak on 16 June 2020 gets re-distributed to the previous days in a suitable way.

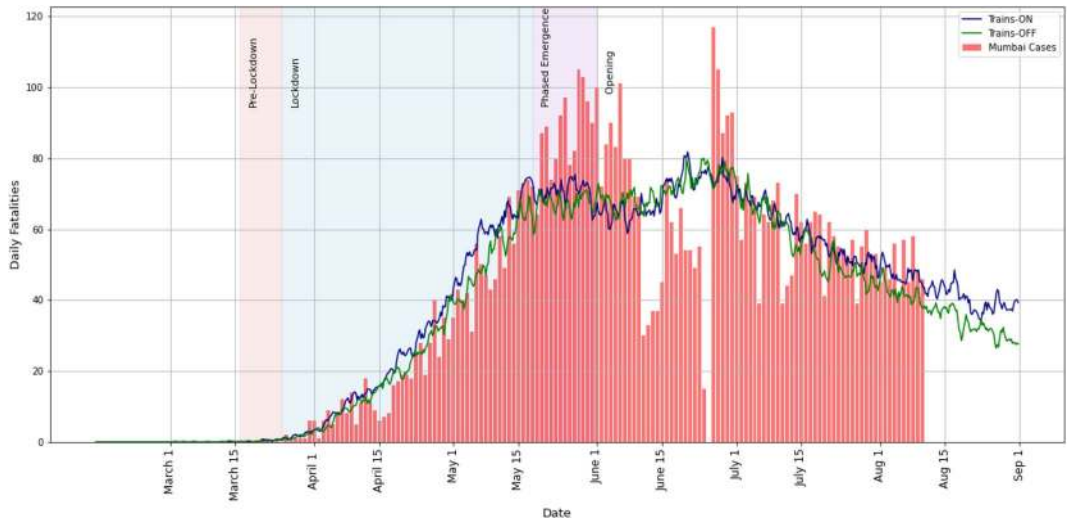
**Table 5:** Simulated Mumbai interventions.

Period	Interventions	Attendance at workplaces	Compliance in non-slums (%)	Compliance in slums (%)
01 March–15 March 2020	No Intervention	at 100% capacity	60	40
16 March–08 April 2020	Lockdown	Essential services operate at 100% capacity, others are closed	60	40
09 April–onwards	Masks ON	Not applicable	60	40
09 April–30 April 2020	Lockdown	Essential services operate at 100% capacity, others are closed	60	40
01 May–17 May 2020	Lockdown with social distancing of the elderly	Essential services operate, others are closed	60	40
18 May–31 May 2020	Social distancing of the elderly, school closure, community_factor=0.75	Essential services operate at 100%, others operate at 5% capacity	60	40
01–30 June 2020	Home quarantine, social distancing of the elderly, school closure, trains ON, community_factor = 0.75	Essential services operate at 100%, others operate at 20% capacity	60	40
01–31 July 2020	Same as above	Essential services operate at 100%, others operate at 33% capacity	60	40
01 August onwards	Same as above	Essential services operate at 100%, others operate at 50% capacity	60	40
Throughout	Soft ward-containment enabled			

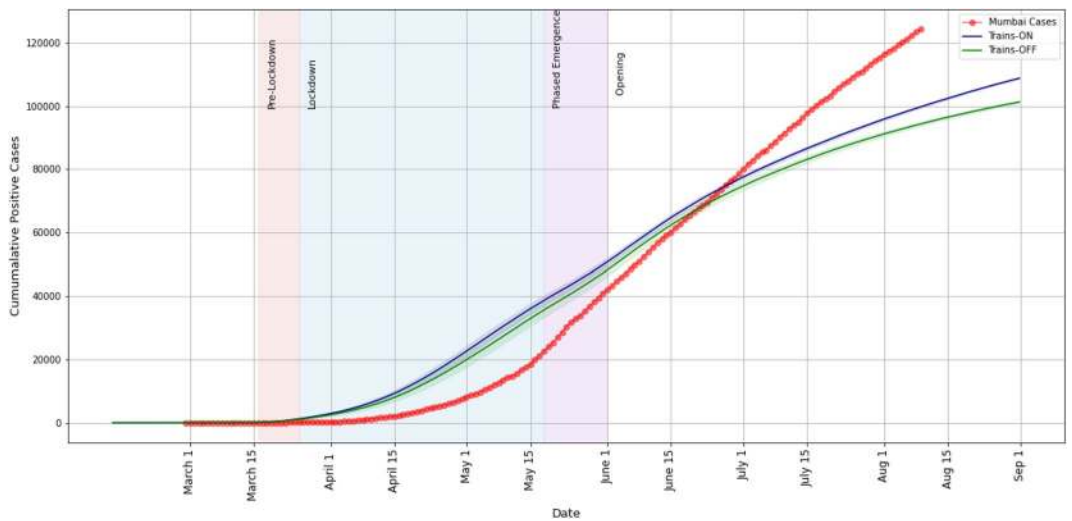
**Figure 19:** Case study C, Sect. 3.3: Mumbai daily cases' estimation.

tation on the testing capacity on the ground. Because of this, the test results of many people arrive late and cases get reported with a certain delay. It is also worth mentioning that,

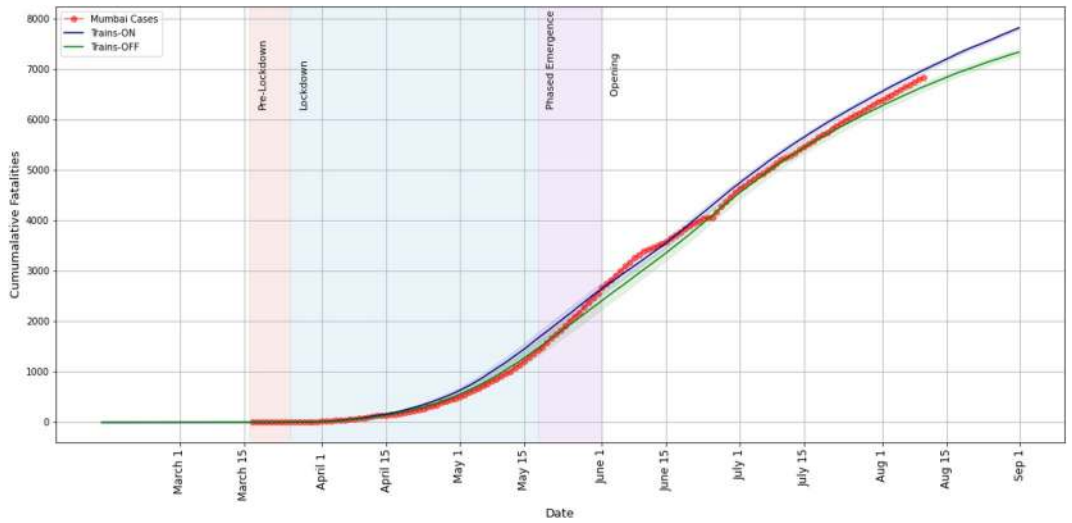
although we over-predict the daily number of cases, we correctly capture the growth rate of the daily number of cases as well as the cumulative number of cases.



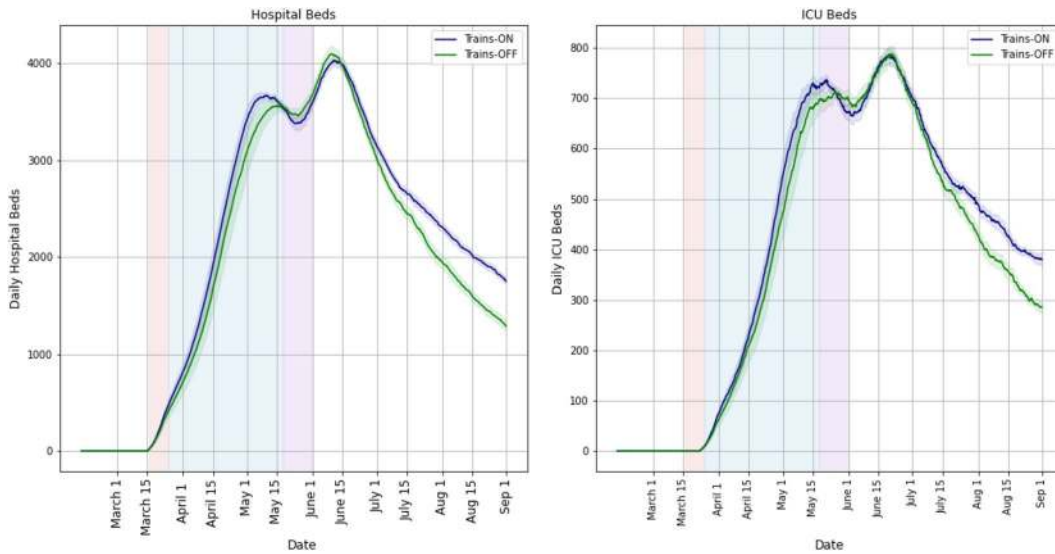
**Figure 20:** Case study C, Sect. 3.3: Mumbai daily fatalities' estimation.



**Figure 21:** Case study C, Sect. 3.3: Mumbai cumulative cases' estimation..



**Figure 22:** Case study C, Sect. 3.3: Mumbai cumulative fatalities' estimation.



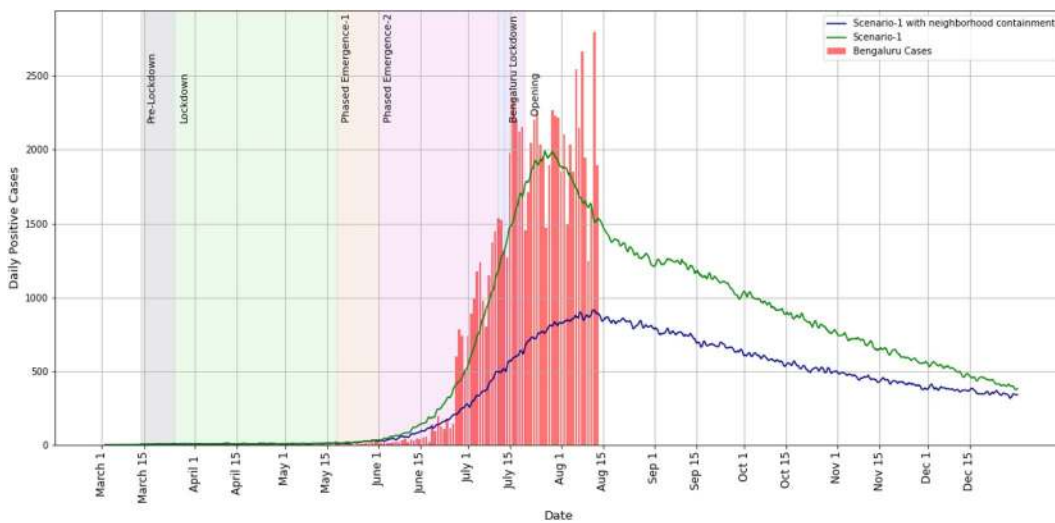
**Figure 23:** Case study C, Sect. 3.3: Mumbai hospital beds' estimation.

### 3.4 Case Study D: Soft Ward-Containment Versus Neighbourhood Containment

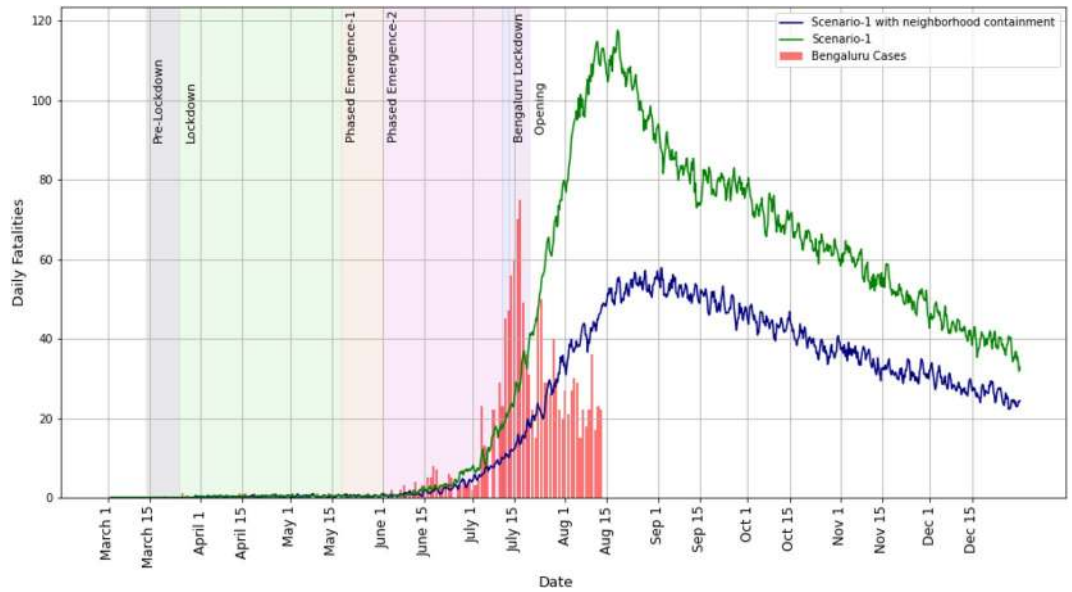
We study the impact of two containment strategies for Bengaluru: soft ward-containment (i.e., linearly varying mobility control that turns an open ward into a locked ward when the number of hospitalised cases becomes 0.1% of the ward's population; in the latter locked scenario, only 25% mobility is allowed for essential services, see Fig. 4) and neighbourhood containment (i.e., when an individual is hospitalised, everyone

living in a 100 m surrounding area undergoes home quarantine). Soft ward-containment is a more feasible strategy than strict ward containment, since the average ward population in Bengaluru is about 62,000. As the number of hospitalised cases in the ward increases, more public-health wardens could be deployed and help to reduce mobility and interaction in the ward.

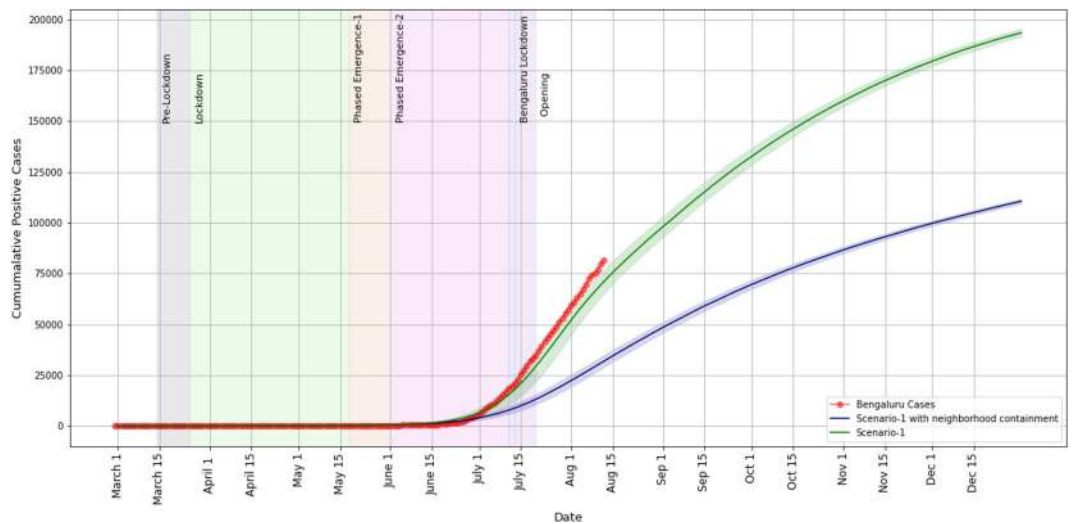
In Figs. 24, 25, 26, 27, and 28, we plot these two scenarios. We observe that neighbourhood containment is more effective than soft ward-containment, in terms of cases and fatalities.



**Figure 24:** Case study D, Sect. 3.4: Bengaluru daily cases' estimation.



**Figure 25:** Case study D, Sect. 3.4: Bengaluru daily fatalities' estimation.



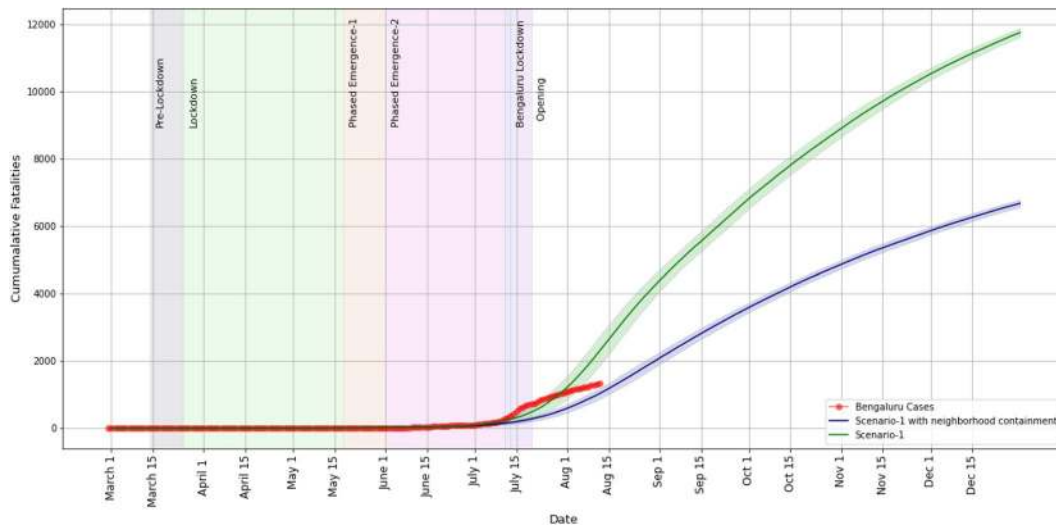
**Figure 26:** Case study D, Sect. 3.4: Bengaluru cumulative cases' estimation.

### 3.5 Case Study E: Soft Ward-Containment at Various Levels in Mumbai

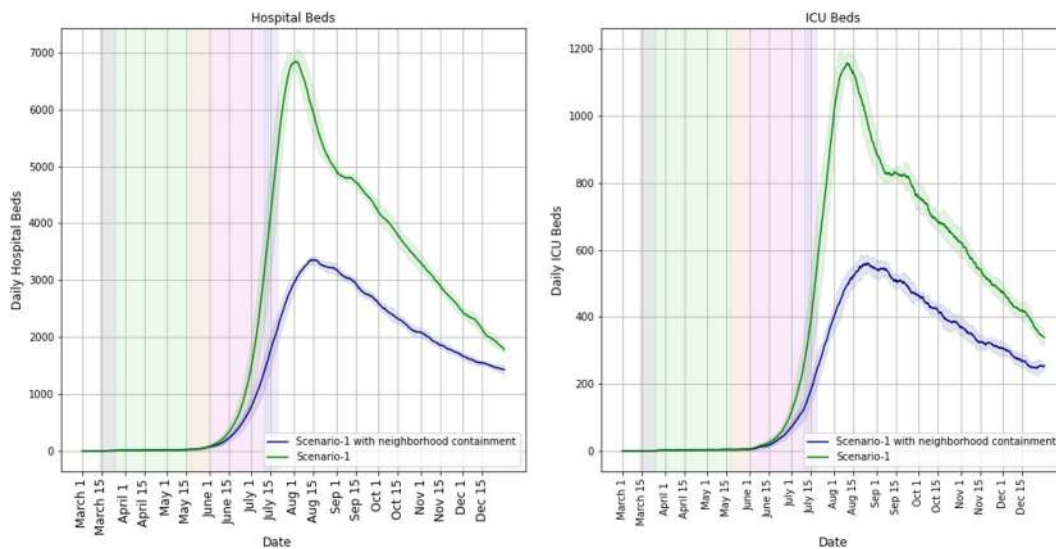
To compare various levels of strictness with which policies are enforced, we now consider the opening scenario indicated in Table 5 and vary the containment leakage to see how this affects the numbers. Containment leakage stands for the level of activity that is allowed in a ward

under containment. A strict enforcement would not allow more than 10% of the normal activity. The case of 'no enforcement' results in activity at 100% of the original level; this corresponds to no adaptation in containment as a function of the number of hospitalisation cases. We explore various values of containment leakage and plot results for 10%, 25%, 50%, and 100%.





**Figure 27:** Case study D, Sect. 3.4: Bengaluru cumulative fatalities' estimation.

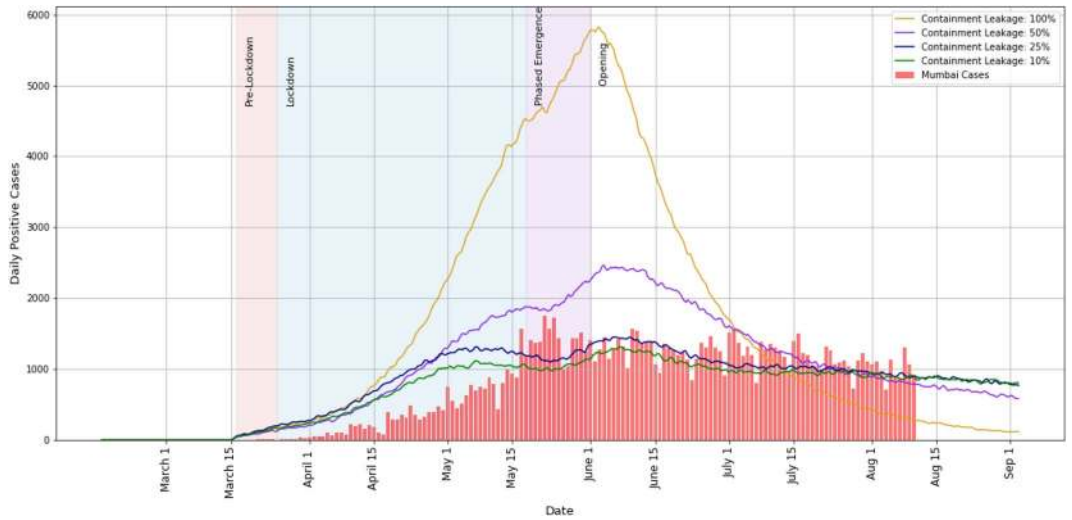


**Figure 28:** Case study D, Sect. 3.4: Bengaluru hospital beds' estimation.

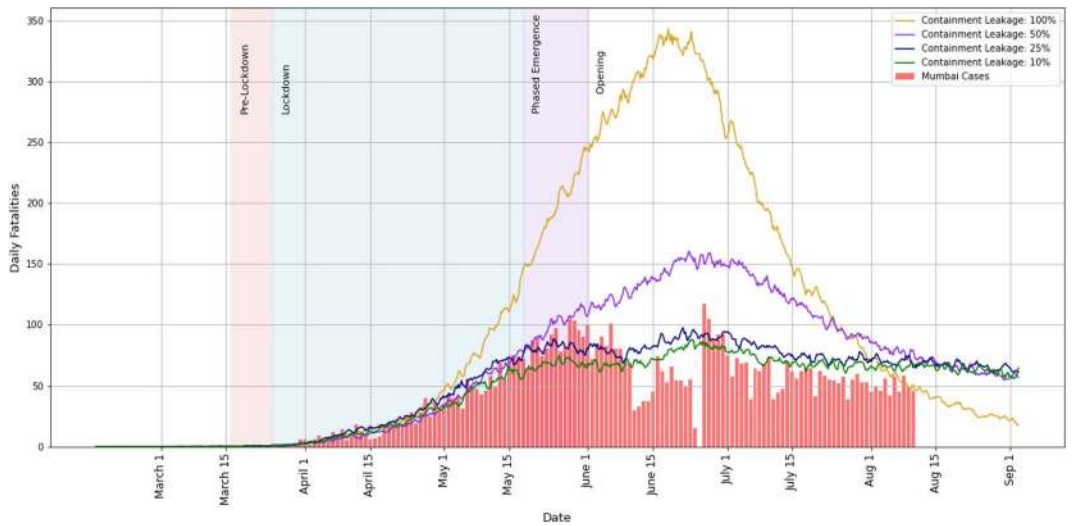
Figures 29, 30, 31, 32, and 33 represent the (simulated) number of daily cases, cumulative cases, daily fatalities, cumulative fatalities, and daily hospital bed estimates, respectively, for the varying containment leakages. These plots demonstrate how an effective containment policy (even as low as 50%) can significantly reduce the number of cases, fatalities, and hospital beds.

An interesting observation is that, for the 25% leakage case, our simulator matches the linear

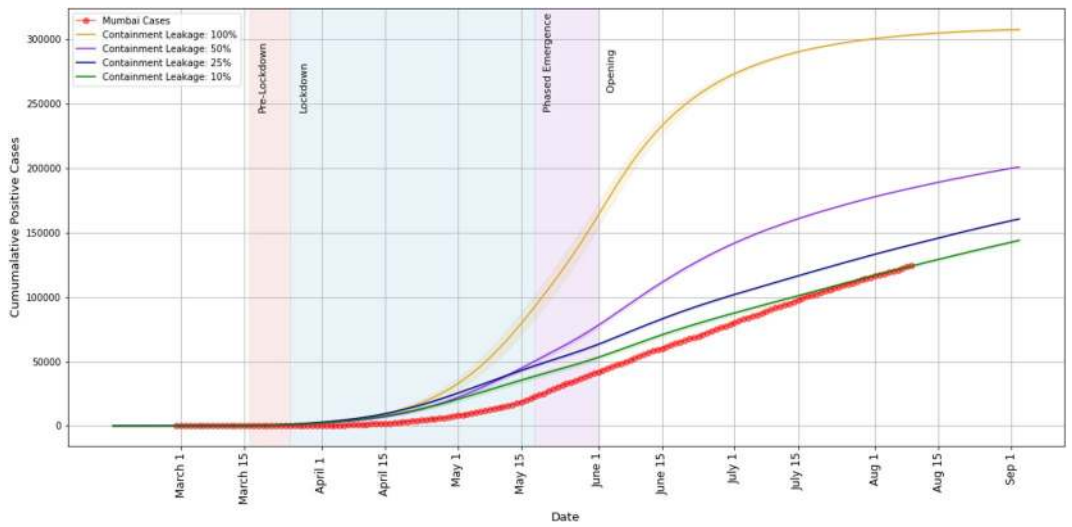
growth trend of the daily and cumulative number of cases as well as fatalities. The linearity is a likely consequence of the specific soft ward-containment policy, triggered by the hospitalisation cases, i.e., mobility is reduced linearly with the number of hospitalisations in that ward. It is not clear to what extent differential equation models can capture such linear trends.



**Figure 29:** Case study E, Sect. 3.5: Mumbai daily cases' estimation.



**Figure 30:** Case study E, Sect. 3.5: Mumbai daily fatalities' estimation.



**Figure 31:** Case study E, Sect. 3.5: Mumbai cumulative cases' estimation.

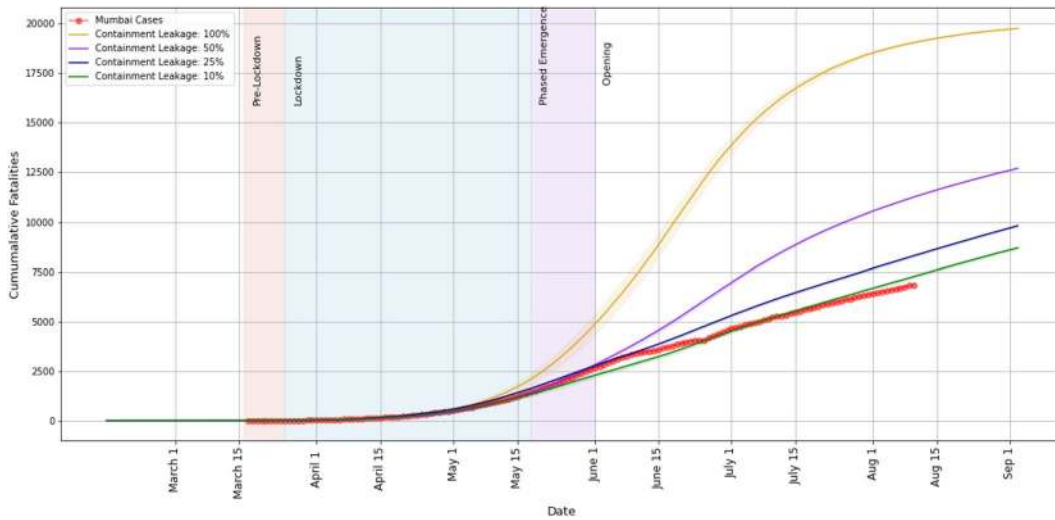


Figure 32: Case study E, Sect. 3.5: Mumbai cumulative fatalities' estimation.

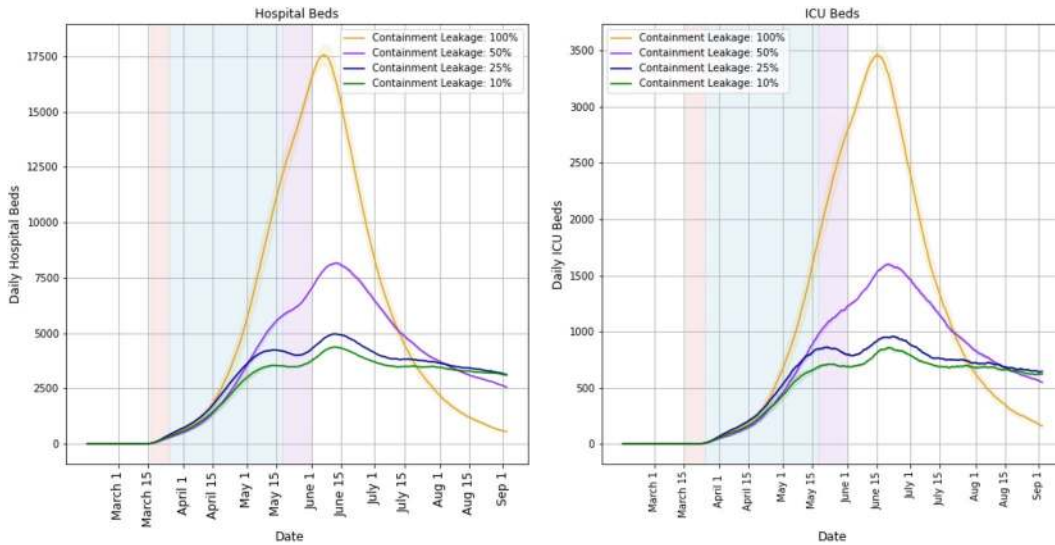


Figure 33: Case study E, Sect. 3.5: Mumbai hospital beds' estimation.

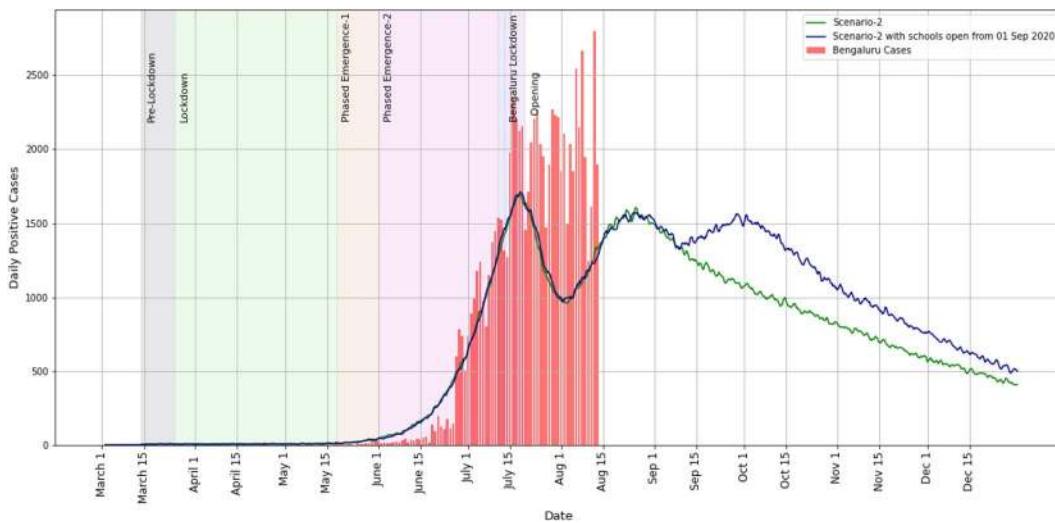
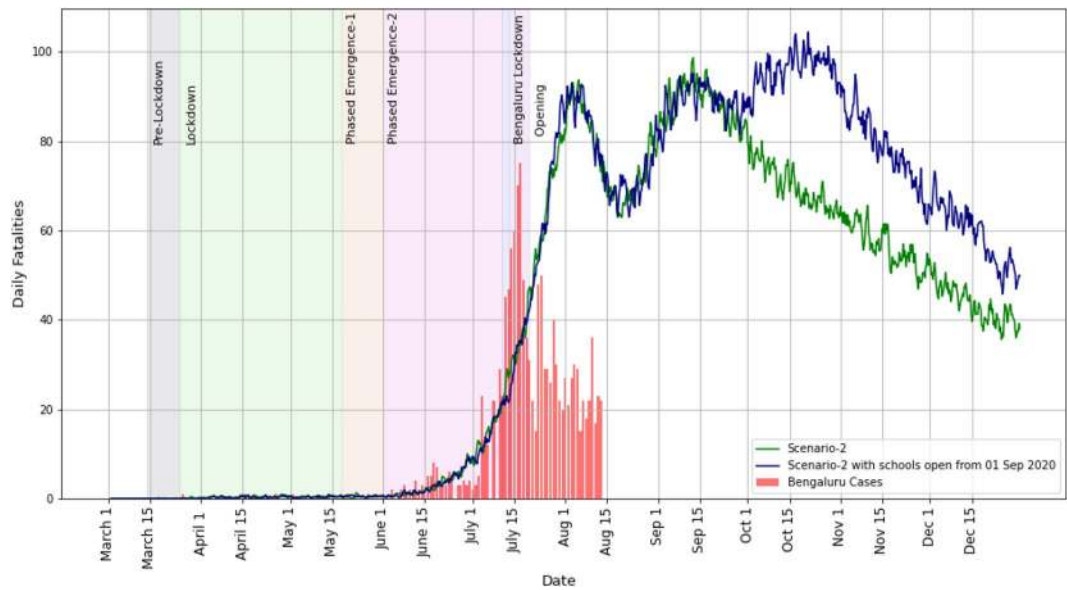
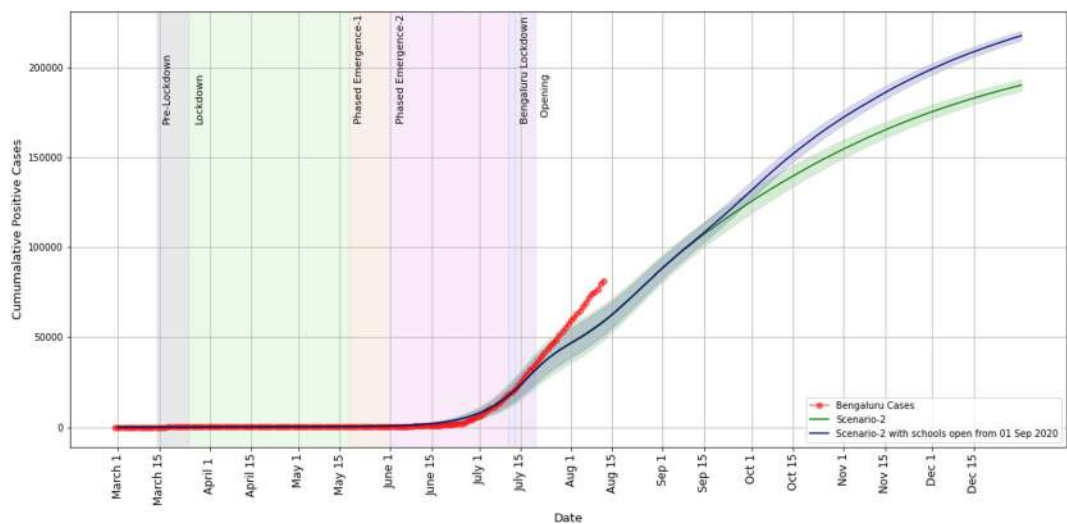


Figure 34: Case study F, Sect. 3.6: Bengaluru daily cases' estimation.



**Figure 35:** Case study F, Sect. 3.6: Bengaluru daily fatalities' estimation.



**Figure 36:** Case study F, Sect. 3.6: Bengaluru cumulative cases' estimation.

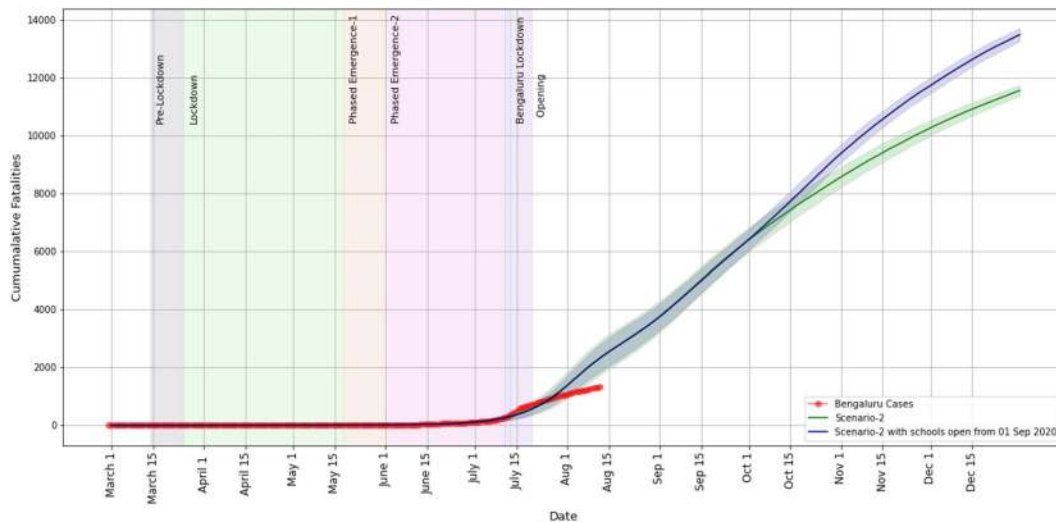
### 3.6 Case Study F: Bengaluru with Schools/Colleges Open from 01 September 2020

We now study the impact of opening schools. In Figs. 34, 35, 36, 37, and 38, we compare the following two scenarios:

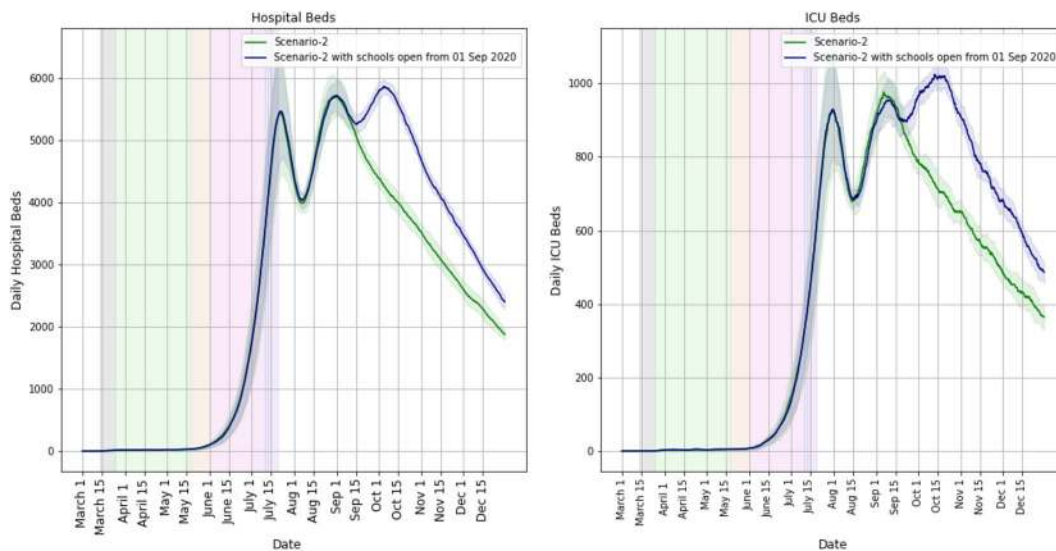
- Schools-closed: The present scenario in Bengaluru, i.e., Scenario-2 in Table 4,
- Schools-open: Scenario-2 in Table 4 with schools open from 01 September 2020.

As expected, both these scenarios follow the same trend until about mid-September, after which the

disease spread increases in the latter. Around early November, we observe a between 10 and 15% increase in the cumulative number of cases and the cumulative number of fatalities due to the opening of schools. This could be weighed with other factors such as the capacity of our healthcare system to handle the rise, the impact of mental health of students due to extended closures, etc., while arriving at a decision on whether schools can be opened from 01 September 2020. The proportion of additional children and adults affected due to the opening of the schools is still under study.



**Figure 37:** Case study F, Sect. 3.6: Bengaluru cumulative fatalities' estimation.



**Figure 38:** Case study F, Sect. 3.6: Bengaluru hospital beds' estimation.

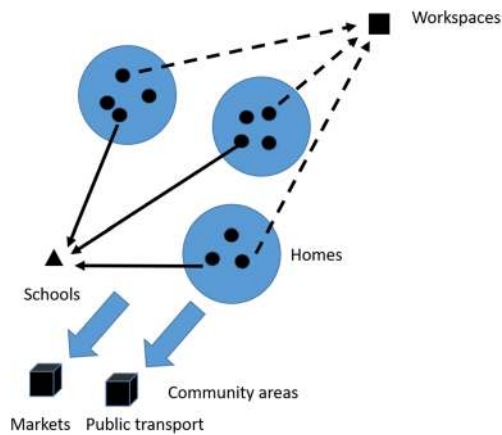
## 4 Simulator

### 4.1 City Generation

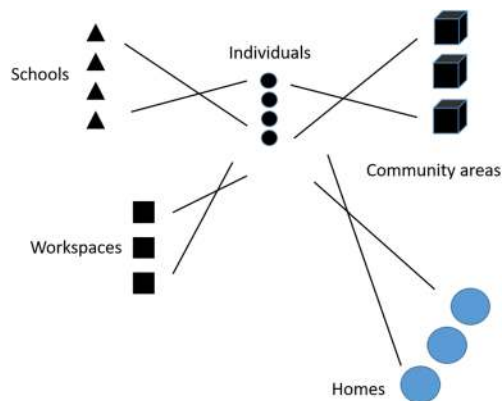
The first step in our agent-based model is to model a synthetic city that respects the demographics of the city that we want to study. Our city generator uses the following data as input:

- Geo-spatial data that provide information on the wards of a city (components) along with boundaries. (If this is not available, one could feed in ward centre locations and ward areas.)
- Population in each ward, with break up on those living in high-density and low-density areas.
- Age distribution in the population.
- Household-size distribution (in high- and low-density areas) and some information on the age composition of the houses (e.g., generation gaps, etc.)
- The number of employed individuals in the city.
- Distribution of the number of students in schools and colleges.
- Distribution of the workplace sizes.





**Figure 39:** Various interaction spaces, solid circles inside homes indicate individuals.



**Figure 40:** Bipartite graph abstraction of interaction spaces.

- Distribution of commute distances.
- Origin-destination densities that quantify movement patterns within the city.

Taking the above data into account, individuals, households, workplaces, schools, transport spaces, and community spaces are instantiated. Individuals are then assigned to households, workplaces or schools, transport, and community spaces, see Fig. 2 for a schematic representation. The algorithms for the assignments do a coarse matching. The matching may be refined as better data become available.

The interaction spaces—households, workplaces or schools, transport, and community spaces—reflect different social networks and transmission happens along their edges. There is interaction among these graphs, because the nodes are common across the graphs, see Fig. 39

for various interaction spaces and Fig. 40 for a bipartite graph abstractions of these interaction spaces. An individual of school-going age who is exposed to the infection at school may expose others at home. This reflects an interaction between the school graph and the household graph. Similarly other graphs interact.

We now describe how individuals are assigned to interaction spaces.

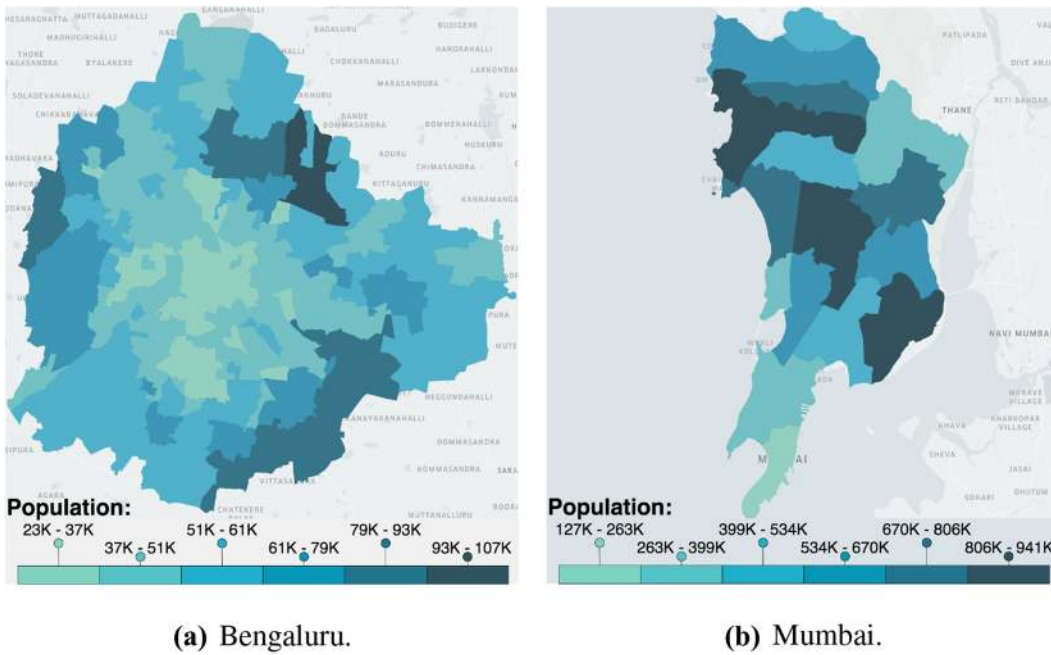
*Individuals and households*  $N$  individuals are instantiated and ages are sampled according to the age distribution in the population. Households (based on the  $N$  and the mean number per household) are then instantiated and assigned a random number of individuals sampled according to the distribution of household sizes. An assignment of individuals to households is then done to match, to the extent possible, the generational structure in typical households. The households are then assigned to wards, so that the total number of individuals in the ward is in proportion to population density in the ward, taken from census data. A population density map is given in Fig. 41a for Bengaluru and in Fig. 41b for Mumbai. The generational gap, household distribution, and age distribution patterns are assumed to be uniform across the wards in the city. Each household in a ward is then assigned a random location in the ward, and all individuals associated with the household are assigned the same geo-location as the household.

Based on the age and the unemployment fraction, each individual is either a student or a worker or neither.

*Assignment of schools* Children of school-going ages 5–14 and a certain fraction of the population aged 15–19 are assigned to schools. These are taken to be students. The remaining fraction of the population aged 15–19 and a certain fraction of the population aged 20–59, based on information on the employed fraction,<sup>8</sup> are all classified as workers and are assigned workplaces. The rest of the population (nonstudent, unemployed) is not assigned to either schools or workplaces.

In past works, given the structure of educational institutions elsewhere, educational institutions have been divided into primary schools, secondary schools, higher secondary schools, and

<sup>8</sup> The unemployed fraction in Bengaluru, from the census data, is just over 50%, even after taking into account employment in the unorganised sector. Similar is the case with Mumbai. This may have some bearing on the epidemic spread.



**Figure 41:** Population density maps of Bengaluru and Mumbai.

universities. The norm in Indian urban areas is that schools handle primary to higher secondary students and then colleges handle undergraduates. We view all such entities as schools.

We assign students to schools on a ward-by-ward basis. In each ward, we have a certain number of students. We pick a school size from a given school-size distribution and instantiate a school of this size and place it randomly in that ward. Students who live in that ward are picked randomly and assigned to this school until that school is filled to its capacity. We repeat this procedure until all students in that ward gets assigned to a school, and then we repeat this procedure for all wards. This procedure could lead to at most one school per ward whose capacity is more than its sampled capacity.

**Assignment of workplaces** Workplace interactions can enable the spread of an epidemic. In principle, Bengaluru's and Mumbai's land-use data could be used to locate office spaces. The assignment of individuals to workplaces is done in two steps. In the first step, for each individual who goes to work, we decide the ward where his/her office is located. This assignment of a "working ward" is based on an Origin–Destination (OD) matrix. An OD matrix is a square matrix whose number of rows equals the number of

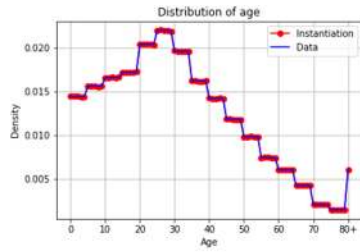
wards, and its  $(i, j)$ th entry tells us the fraction of people who travel from ward  $i$  to ward  $j$  for work. In the second step, for each ward, we sample a workplace size from a workplace size distribution, create a workplace of this capacity, and place it uniformly-at-random in that ward. We then randomly assign individuals who work in that ward to this workplace. Similar to assignment of schools, we continue to create workplaces in this ward until every individual working in that ward gets assigned to a workplace, and we repeat this procedure for all wards. For Bengaluru, the OD matrix is obtained from the regional travel model used for Bengaluru. For Mumbai, based on the "zone to zone" travel data from<sup>21</sup>, an origin–destination matrix was extrapolated based on the population of each ward.

The above assignments could be improved further in later versions of this simulator.

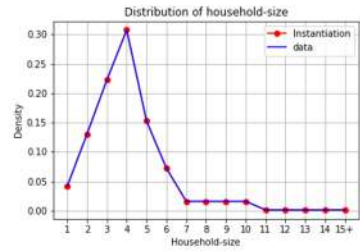
**Community spaces** Community spaces include day care centres, clinics, hospitals, shops, markets, banks, movie halls, marriage halls, malls, eateries, food messes, dining areas and restaurants, public transit entities like bus stops, metro stops, bus terminals, train stations, airports, etc. While we hope to return to model a few of the important ones explicitly at a later time, we proceed along the route taken by<sup>22</sup> with two modifications.

In our current implementation, each individual sees one community that is personalised to the individual's location and age and one

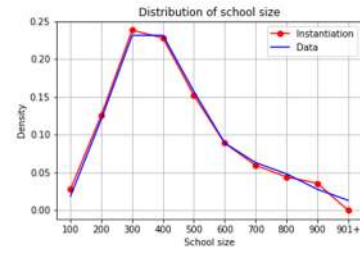
<sup>9</sup> This dimension 178 m comes from an approximate "squaring of a circle" of 100 m radius



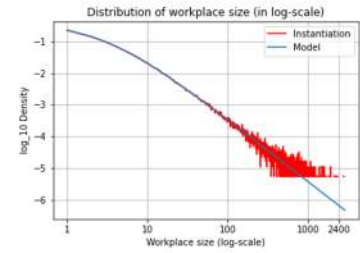
(a) Bengaluru – Age distribution



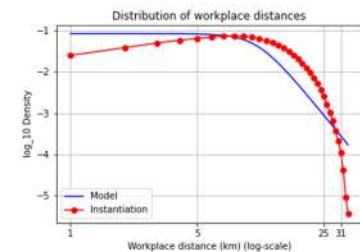
(b) Bengaluru – Household size distribution



(c) Bengaluru – School size distribution



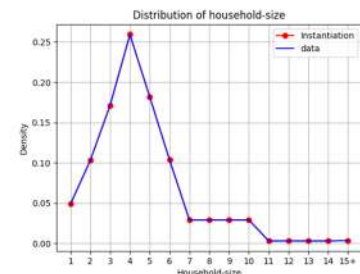
(d) Bengaluru – Workplace size distribution



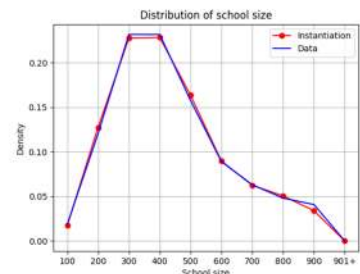
(e) Bengaluru – Commuter distance distribution



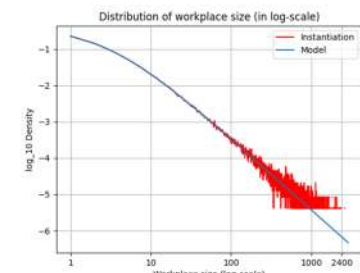
(f) Mumbai – Age distribution



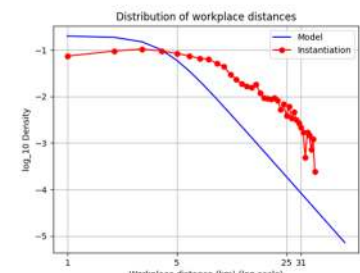
(g) Mumbai – Household size distribution



(h) Mumbai – School size distribution

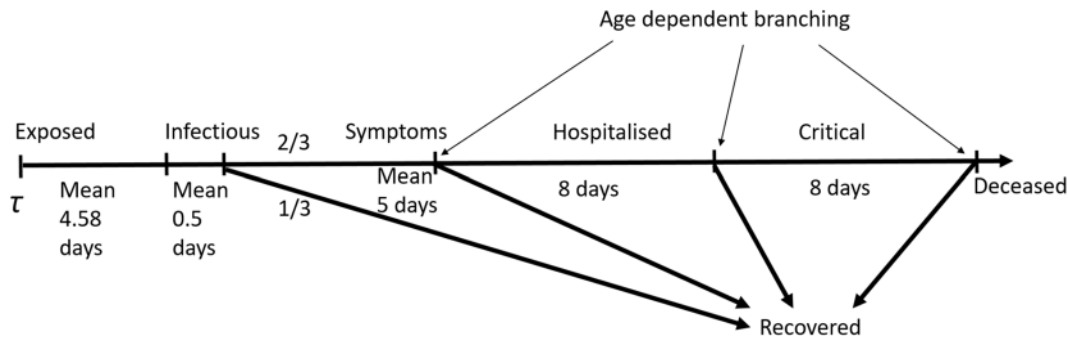


(i) Mumbai – Workplace size distribution



(j) Mumbai – Commuter distance distribution

**Figure 42:** Validation of our synthetic Bengaluru and Mumbai. **a–e** Shows the validation plots for Bengaluru and **f–j** shows the validation plots for Mumbai.



**Figure 43:** A simplified model of COVID-19 progression.

**Table 6:** Hospitalisation estimates taken from<sup>6</sup>, based on studies in<sup>25</sup>.

Age group (years)	% Symptomatic cases Requiring hospitalisation	% Hospitalised cases Requiring critical care	% Critical cases Deceased
0–9	0.1	5.0	40
10–19	0.3	5.0	40
20–29	1.2	5.0	50
30–39	3.2	05.0	50
40–49	4.9	6.3	50
50–59	10.2	12.2	50
60–69	16.6	27.4	50
70–79	24.3	43.2	50
80+	27.3	70.9	50

transport space personalised to the individual’s commute distance. For ease of implementation, the personalisation of the community space is based on ward-level common local communities and a distance-kernel-based weighting. The personalisation of the transport space is based on commute distance. Details are given in Sect. 4.3.

*Age-stratified interaction* The interactions across these communities could be age-stratified. This may be informed by social networks studies, for, e.g., as in<sup>23</sup> which has been used in a recent compartmentalised SEIR model<sup>24</sup>.

*Smaller subnetworks* We create smaller subnetworks in workplaces, schools, and communities, and associate certain number people to these smaller networks with the interpretation that people in a smaller subnetwork have high contact rate among them compared to the others. In some more detail, we create “project” networks at each workplace consisting of people in that workplace having closer interaction, a “class” network in each school consisting of students of the same age, a random community network among people in a given ward to model daily random interactions, and a neighbourhood subnetwork among people living in a 178 m × 178 m square.<sup>9</sup>

These subnetworks are later used for identifying and testing/quarantining individuals based on a contact tracing protocol.

The output of all the above is our synthetic city on which infection spreads. Figure 42 provides an indication of how close our synthetic city is to the true city in terms of the indicated statistics.

**4.2 Disease Progression**

We have used a simplified model of COVID-19 progression, based on descriptions in<sup>25</sup> and<sup>6</sup>. This will need updating as we get India specific data.

An individual may have one of the following states, see Fig. 43: susceptible, exposed, infective (pre-symptomatic or asymptomatic), recovered, symptomatic, hospitalised, critical, or deceased.

We assume that, initially, the entire population is susceptible to the infection. Let  $\tau$  denote the time at which an individual is exposed to the

<sup>10</sup> This needs to be updated based on hospitalisation guidelines.

<sup>11</sup> It is possible to add comorbidities—diabetes, hypertension, etc.—in addition to age. Mortality and prognosis appear to depend heavily on comorbidities. We leave it for the future.

virus, see Fig. 43. The incubation period is random with the Gamma distribution of shape 2 and scale 2.29; the mean incubation period is then 4.58 days (4.6 days in<sup>6</sup> and 4.58 in<sup>26</sup>). Individuals are infectious for an exponentially distributed period of mean duration 0.5 of a day. This covers both pre-symptomatic transmission and possible asymptomatic transmission. We assume that a third of the patients recover, these are the asymptomatic patients; the remaining two-third develop symptoms. Estimates of the number of asymptomatic patients vary from 0.2 to 0.6. Though we have explored other asymptomatic fractions, we restrict attention here to 1/3. Symptomatic patients are assumed to be 1.5 times more infectious during the symptomatic period than during the pre-symptomatic but infective stage. Individuals either recover or move to the hospital after a random duration that is exponentially distributed with a mean of 5 days.<sup>10</sup> The probability that an individual recovers depends on the individual's age.<sup>11</sup> It is also assumed that recovered individuals are no longer infective nor susceptible to a second infection. While hospitalised individuals may continue to be infectious, they are assumed to be sufficiently isolated, and hence do not further contribute to the spread of the infection. Further progression of hospitalised individuals to critical care is mainly for assessing the need for hospital beds, intensive care unit (ICU) beds, critical care equipments, etc. This will need to be adapted to our local hospital protocol.

Let us reiterate. Once a susceptible individual has been exposed, the trajectory in Fig. 43 takes over for that individual. Further progressions are (in our current implementation) only based on the agent's age.

### 4.3 Disease Spread

At each time  $t$ , an infection rate  $\lambda_n(t)$  is computed for each individual  $n$  based on the prevailing conditions. In the time duration  $\Delta t$  following time  $t$ , each susceptible individual moves to the exposed state with probability  $1 - \exp\{-\lambda_n(t) \cdot \Delta t\}$ , independently of all other events. Other transitions are as per the disease progression described earlier. Time is then updated to  $t + \Delta t$ , and the conditions are then updated to reflect the new exposures, changes to infectiousness, hospitalisations, recoveries, contact tracing, quarantines, tests, test outcomes, etc., during the period  $t$  to  $t + \Delta t$ . The process outlined at the beginning of this paragraph is repeated until the end of the simulation.  $\Delta t$  was taken to be 6 h in our simulator and is configurable (Table 6).

A city of  $N$  individuals,  $H$  households,  $S$  schools,  $W$  workplaces, one community space (comprising  $C$  wards), one transport space, and associations of individuals to these entities is the starting point for the infection spread simulator. Infection spread is then implemented as follows.

An individual  $n$  can transmit the virus in the infective (pre-symptomatic or asymptomatic stage) or in the symptomatic stage. At time  $t$ , this is indicated as  $I_n(t) = 1$  when infective and otherwise  $I_n(t) = 0$  otherwise.

Additionally, each individual has two other parameters: a severity variable  $C_n$  and a relative infectiousness variable  $\rho_n$ , see<sup>22</sup>. Both bring in heterogeneity to the model. Severity  $C_n = 1$  if the individual suffers from a severe infection and  $C_n = 0$  otherwise; this is sampled at 50% probability independently of all other events. Infectiousness  $\rho_n$  is a random variable that is Gamma distributed with shape 0.25 and scale 4 (so the mean is 1). The severity variable captures severity-related absenteeism at school/workplace, associated decrease of infection spread at school/workplace, and the increase of infection spread at home.

If the individual gets exposed at time  $\tau_n$ , a relative infection-stage-related infectiousness is taken to be  $\kappa(t - \tau_n)$  at time  $t$ . For the disease progression described in the previous section, this is 1 in the pre-symptomatic and asymptomatic stages, 1.5 in the symptomatic, hospitalised, and critical stages, and 0 in the other stages.

To describe the infection spread at transport spaces, let  $\mathcal{T}(n) = 1$  if agent  $n$  uses public transport and let  $\mathcal{T}(n) = 0$  otherwise. Let  $A_{n,t} = 0$  if, at time  $t$ , agent  $n$  is either (i) compliant and under quarantine, (ii) hospitalised, (iii) critical, or (iv) dead, and let  $A_{n,t} = 1$  if none of the above is true and agent  $n$  attends office at time  $t$ . We model the effectiveness of masks by reducing the ability of an infectious individual to transmit the infection by 20% if a mask is worn (see<sup>13-16</sup>); let  $M_n = 0.8$  if agent  $n$  wears a face mask in public transport and  $M_n = 1$  otherwise.

Let  $\beta_h, \beta_s, \beta_w, \beta_T, \beta_c, \beta_h^*, \beta_s^*, \beta_w^*,$  and  $\beta_c^*$  denote the transmission coefficients at home, school, workplace, transport, community spaces, neighbourhood network, class network, project network, and random community network, respectively. These can be viewed as scaled contact rates with members in the household, school, workplace, community, neighbourhood, class, project, and random community, respectively. More precisely, these are the *expected number of eventful (infection spreading) contact opportunities* in each of these interaction spaces. It accounts for



the combined effect of frequency of meetings and the probability of infection spread during each meeting.

For a susceptible individual, the rate of transmission is governed by the sum of product of contact rate  $\beta$  and infectiousness in all the interactions spaces. To model infectiousness, we consider three scenarios.

*Interactions without age stratification:* This is the simplest model where interactions within each network are assumed to be homogeneous. A susceptible individual  $n$  (who belongs to home  $h(n)$ , school  $s(n)$ , workplace  $w(n)$ , transport space  $T(n)$ , and community space  $c(n)$ ) sees the following infection rate at time  $t$ :

$$\begin{aligned} \lambda_n(t) = & \sum_{n':h(n')=h(n)} \frac{1}{n_{h(n)}^\alpha} \cdot I_{n'}(t)\beta_h\kappa(t - \tau_{n'})\rho_{n'}(1 + C_{n'}(\omega - 1)) \\ & + \sum_{n':s(n')=s(n)} \frac{1}{n_{s(n)}} \cdot I_{n'}(t)\beta_s\kappa(t - \tau_{n'})\rho_{n'}(1 + C_{n'}(\omega\psi_s(t - \tau_n) - 1)) \\ & + \sum_{n':w(n')=w(n)} \frac{1}{n_{w(n)}} \cdot I_{n'}(t)\beta_w\kappa(t - \tau_{n'})\rho_{n'}(1 + C_{n'}(\omega\psi_w(t - \tau_n) - 1)) \\ & + \frac{\sum_{n':T(n')=1} A_{n',t}}{\sum_{n'} T(n')} \times \sum_{n':T(n')=T(n)} \left( \frac{d_{n',w(n')}I_{n'}(t)\beta_T M_{n'}}{\sum_{n':T(n')=T(n)} d_{n',w(n')}} \right) \\ & + \frac{\zeta(a_n) \cdot f(d_{n,c})}{\sum_{c'} f(d_{c,c'})} \sum_{c'} f(d_{c,c'})h_{c,c'}(t), \end{aligned} \tag{1}$$

where

mission rate from individuals in ward  $c'$  to an

$$h_{c,c'}(t) = \left( \frac{\sum_{n':c(n')=c'} f(d_{n',c(n')}) \cdot \zeta(a_{n'}) \cdot I_{n'}(t)\beta_c r_c \kappa(t - \tau_{n'})\rho_{n'}(1 + C_{n'}(\omega - 1))}{\sum_{n'} f(d_{n',c(n')})} \right). \tag{2}$$

The expression (1) can be viewed as the rate at which the susceptible individual  $n$  contracts the infection at time  $t$ . Each of the components on the right-hand side indicates the rate from home, school, workplace, transport space, and community. The additional quantities, over and above what we have already described, are as follows.

The parameter  $\alpha$  determines how household transmission rate scales with household size, a crowding-at-household factor. It increases the propensity to spread the infection by a factor  $n^{1-\alpha}$ . We have taken  $\alpha = 0.8$ , see<sup>22</sup>.

A common parameter  $\omega$  indicates how a severely infected person affects a susceptible one, as will be clear from below. (This is to be tuned at a later stage and is set to 2 now.)

The functions  $\psi_s(\cdot)$  and  $\psi_w(\cdot)$  account for absenteeism in case of a severe infection. It can

be time-varying and can depend on school or workplace. We take  $\psi_s(t) = 0.1$  and  $\psi_w(t) = 0.5$  while infective and after 1 day since infectiousness. School goes with severe infection contribute lesser to the infection spread, due to higher absenteeism, than those that go to workplaces; moreover, the absenteeism results in an increased spreading rate at home.

The function  $\zeta(a)$  is the relative travel-related contact rate of an individual aged  $a$ . We take this to be 0.1, 0.25, 0.5, 0.75, 1, 1, 1, 1, 1, 1, 0.75, 0.5, 0.25, and 0.1 for the various age groups in steps of 5 years, with the last one being the 80+ category.

The quantity  $h_{c,c'}(t)$  represents the trans-

individual in ward  $c$ . As above, each individual contributes in a distance-weighted way in how an individual in a ward  $c'$  affects another individual in another ward  $c$ .

The factor  $r_c$  stands for a high-density interaction multiplying factor. For Mumbai,  $r_c = 2$  for some high-density areas and  $r_c = 1$  for the other areas. For Bengaluru,  $r_c = 1$  for all wards.

The function  $f(\cdot)$  is a distance kernel that can be matched to the travel patterns in the city.

Finally, our choice of the infection rate from the community space is a little different from the rate specified in<sup>22</sup>, to enable an efficient implementation. When the distance kernel is  $f(d) = 1/(1 + (d/a)^b)$  and  $d \ll a$ , i.e., the wards are small, then our specification is close to

**Table 7:** Model parameters.

Parameter	Symbol	Bengaluru	Mumbai
Transmission coefficient at home	$\beta_h$	1.0884 (calibrated)	0.7928 (calibrated)
Transmission coefficient at school	$\beta_s$	0.2548 (calibrated)	0.2834 (calibrated)
Transmission coefficient at workplace	$\beta_w$	0.1274 (calibrated)	0.1417 (calibrated)
Transmission coefficient at community	$\beta_c$	0.0169 (calibrated)	0.0149 (calibrated)
Subnetwork upscale factor	$\tilde{\beta}$	9	9
Transmission coefficient at transport space	$\beta_t$	0	0.1506
Household crowding	$1 - \alpha$	0.2	0.2
Community crowding	$r_c$	1	2
Distance kernel $f(d) = 1/(1 + (d/a)^b)$	$(a, b)$	(10.751, 5.384)	(2.709, 1.279)
Infectiousness shape (Gamma distributed)	(shape, scale)	(0.25, 4)	(0.25, 4)
Severity probability	$\Pr\{C_n = 1\}$	0.5	0.5
Age stratification	$M_{n,n'}$	Not used	Not used
Project subnetwork size range	$n_{\mathcal{W}(n)}$	3 – 10	3 – 10
Family friends' subnetwork range	no symbol	2-5 families	2-5 families

that indicated in<sup>22</sup>. We take  $a = 10.751$  km and  $b = 5.384$ , based on a fit on data for Bengaluru.

As one can see from (1), we have one community space but with contributions from various wards. This enables inclusion of 'containment zones' and the associated restriction of interaction across such zones, as we shall soon describe.

*Age-stratified interactions* If this is enabled, the home, school, workplace, and community interaction rates have an extra factor  $M_{n,n'}^h, M_{n,n'}^s$ , and  $M_{n,n'}^w$  in the summand which accounts for age-stratified interactions. Each of these depends on  $n$  and  $n'$  only through the ages of agents  $n$  and  $n'$ . The resulting contact rate for individual  $n$  at time  $t$  is then:

where  $h_{c,c'}(t)$  is given in (2). Computational complexity can be reduced by focusing only on the principal components of  $M^h, M^s$ , and  $M^w$ .

*Interactions with smaller subnetworks* In this situation, we have additional contact rate parameters, one for each smaller subnetwork: let  $\beta_h^*, \beta_s^*, \beta_w^*$  and  $\beta_c^*$  denote the transmission coefficients at neighbourhood network, class network, project network, and random community network, respectively. Then, an agent  $n$  [who belongs to neighbourhood network  $\mathcal{H}(n)$ , class  $\mathcal{S}(n)$ , project  $\mathcal{W}(n)$ , and random community  $\mathcal{C}(n)$ , in addition to home  $h(n)$ , school  $s(n)$ , workplace

$$\begin{aligned}
 \lambda_n(t) = & \sum_{n':h(n')=h(n)} \frac{M_{n,n'}^h}{n_{h(n)}^\alpha} \cdot I_{n'}(t) \beta_h \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega - 1)) \\
 & + \sum_{n':s(n')=s(n)} \frac{M_{n,n'}^s}{n_{s(n)}} \cdot I_{n'}(t) \beta_s \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega \psi_s(t - \tau_n) - 1)) \\
 & + \sum_{n':w(n')=w(n)} \frac{M_{n,n'}^w}{n_{w(n)}} \cdot I_{n'}(t) \beta_w \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega \psi_w(t - \tau_n) - 1)) \\
 & + \frac{\sum_{n':T(n')=1} A_{n',t}}{\sum_{n'} T(n')} \times \sum_{n':T(n')=T(n)} \left( \frac{d_{n',w(n')} I_{n'}(t) \beta_T M_{n'}}{\sum_{n':T(n')=T(n)} d_{n',w(n')}} \right) \\
 & + \frac{\zeta(a_n) \cdot f(d_{n,c})}{\sum_{c'} f(d_{c,c'})} \sum_{c'} f(d_{c,c'}) h_{c,c'}(t),
 \end{aligned} \tag{3}$$

$w(n)$ , transport space  $T(n)$ , and community space  $c(n)$ ] sees the following infection rate at time  $t$ :

$$\begin{aligned}
 \lambda_n(t) = & \sum_{n':h(n')=h(n)} \frac{1}{n_{h(n)}^\alpha} \cdot I_{n'}(t) \beta_h \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega - 1)) \\
 & + \zeta(a_n) \sum_{n':\mathcal{H}(n')=\mathcal{H}(n)} \frac{1}{n_{\mathcal{H}(n)}} \cdot \zeta(a_{n'}) I_{n'}(t) \beta_h^* \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega - 1)) \\
 & \text{(larger neighbourhood interaction)} \\
 & + \frac{\zeta(a_n) f(d_{n,c(n)})}{\sum_{n':\mathcal{C}(n')=\mathcal{C}(n)} f(d_{n',c(n')})} \\
 & \quad \times \sum_{n':\mathcal{C}(n')=\mathcal{C}(n)} f(d_{n',c(n')}) \zeta(a_{n'}) I_{n'}(t) \beta_c^* \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega - 1)) \\
 & \text{(close friends' circle interaction)} \\
 & + \sum_{n':s(n')=s(n)} \frac{1}{n_{s(n)}} \cdot I_{n'}(t) \beta_s \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega \psi_s(t - \tau_n) - 1)) \\
 & + \sum_{n':\mathcal{S}(n')=\mathcal{S}(n)} \frac{1}{n_{\mathcal{S}(n)}} \cdot I_{n'}(t) \beta_s^* \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega \psi_s(t - \tau_n) - 1)) \\
 & \text{(class network interaction)} \\
 & + \sum_{n':w(n')=w(n)} \frac{1}{n_{w(n)}} \cdot I_{n'}(t) \beta_w \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega \psi_w(t - \tau_n) - 1)) \\
 & + \sum_{n':\mathcal{W}(n')=\mathcal{W}(n)} \frac{1}{n_{\mathcal{W}(n)}} \cdot I_{n'}(t) \beta_w^* \kappa(t - \tau_{n'}) \rho_{n'} (1 + C_{n'}(\omega \psi_w(t - \tau_n) - 1)) \\
 & \text{(project network interaction)} \\
 & + \frac{\sum_{n':T(n')=1} A_{n',t}}{\sum_{n'} T(n')} \times \sum_{n':T(n')=T(n)} \left( \frac{d_{n',w(n')} I_{n'}(t) \beta_T M_{n'}}{\sum_{n':T(n')=T(n)} d_{n',w(n')}} \right) \\
 & + \frac{\zeta(a_n) \cdot f(d_{n,c})}{\sum_{c'} f(d_{c,c'})} \sum_{c'} f(d_{c,c'}) h_{c,c'}(t),
 \end{aligned} \tag{4}$$

where  $h_{c,c'}(t)$  is given in (2). The subnetwork interactions are stronger contexts for disease spread. Contact tracing targets exactly these subnetworks for additional testing, case isolation, or quarantine.

#### 4.4 Seeding of Infection

Two methods of seeding the infection have been implemented.

- A small number of individuals can be set to either exposed, pre-symptomatic/asymptomatic, or symptomatic states, at time  $t = 0$ , to seed the infection. This can be done randomly based either on ward-level probabilities, which could be input to the simulator, or it can be done uniformly-at-random across all wards in the city.

in the first stage of infectiousness (pre-symptomatic or asymptomatic). This could be done based on data for patients with a foreign travel history who eventually visited a hospital. A certain multiplication factor then accounts for the asymptomatic and the symptomatic individuals that recover without the need to visit the hospital. The seeding is done at a random time earlier in the time line, based on the disease progression.

#### 4.5 Calibration

We calibrate our model by tuning the transmission coefficients at various interaction spaces under the no-intervention scenario to match the cumulative fatalities to a target curve. We assume a common upscaling factor  $\tilde{\beta}$  for the transmission coefficients of smaller subnetworks, i.e., we

**Table 8:** Interventions.

Label	Policy	Description
CI	Case isolation at home	Symptomatic individuals stay at home for 7 days, non-household contacts reduced by 90% during this period, household contacts reduce by 25%
HQ	Voluntary home quarantine	Once a symptomatic individual has been identified, all members of the household remain at home for 14 days. Non-household contacts reduced by 90% during this period, household contacts reduce by 25%
SDO	Social distancing of those aged 65 and over	Non-household contacts reduce by 75%
LD	Lockdown	Closure of schools and colleges. Only essential workplaces active. For a compliant household, household contact rate doubles, community contact rate reduces by 75%, workspace contact rate reduces by 75%. For a non-compliant household, household contact rate increases by 25%, workspace contact rate reduces by 75%, and no change to community contact rate
LD40-CI	Lockdown for 40 days	Lockdown for 40 days and then normal activity, but with CI
LD40-PE-CI	One particular phased emergence (PE) from lockdown	Lockdown for 40 days, then CI, HQ and SDO for 14 days. Schools and colleges remain closed during this period. Normal activity resumes after this period with reopening of schools and colleges, but with CI
LD40-PE-SCCI	Another phased emergence from lockdown	Lockdown for 40 days, then CI, HQ, and SDO for 14 days. Schools and colleges remain closed during this period (SC). Normal activity resumes after this period, but schools and colleges remain closed for another 28 days (SC). CI remains in place throughout
LD40-PEOE-CI	A third type of phased emergence from lockdown	Lockdown for 40 days, and then CI, HQ, and SDO for 14 days. Schools and colleges remain closed and an odd-even workplace strategy is in place during this period. Normal activity resumes after this period. CI remains in force throughout

set  $\beta_w^* = \tilde{\beta}\beta_w$ ,  $\beta_s^* = \tilde{\beta}\beta_s$ , and  $\beta_h^* = \beta_c^* = \tilde{\beta}\beta_c$ . We assume that  $\tilde{\beta} = 9$ , indicating that the sub-networks account for 90% of the overall contacts. The following heuristic iterative algorithm inspired by stochastic approximation is then used to identify the best choice of the free parameters:

stochasticity from the simulator, in each step  $n$ , we run the simulator 6 times and take the average values of  $\Lambda_h(n)$ ,  $\Lambda_w(n)$ ,  $\Lambda_c(n)$  and  $m(n)$ . We stop the algorithm at time  $n$  if we meet our targets, i.e., if:

$$\beta_h(n+1) = \left( \beta_h(n) - \frac{\Lambda_h(n)}{n+3} \right) \times [\exp(m^* - m(n))]^{1/a},$$

$$\beta_w(n+1) = \left( \beta_w(n) - \frac{\Lambda_w(n)}{n+3} \right) \times [\exp(m^* - m(n))]^{1/a},$$

$$\beta_c(n+1) = \left( \beta_c(n) - \frac{\Lambda_c(n)}{n+3} \right) \times [\exp(m^* - m(n))]^{1/a},$$

where  $[\exp(m^* - m(n))]^{1/a} = \min\{\max\{\exp(m^* - m(n)), a\}, 1/a\}$ ,  $\Lambda_h(n)$  (resp.  $\Lambda_w(n)$ ,  $\Lambda_c(n)$ ) is the fraction of infections from home (resp. workplace, community) in the  $n$ th step, and  $m(n)$  is the slope of the cumulative fatalities curve in log scale in an initial linear region, obtained by running the simulator on a smaller city file (of 1 million population) in the no-intervention scenario. We set  $a = 2/3$ . To minimise the effect of

$$\begin{aligned} |m(n) - m^*| &\leq 0.01, \\ |\Lambda_h(n) - \Lambda_h^*| &\leq 0.01, \\ |\Lambda_w(n) - \Lambda_w^*| &\leq 0.01, \text{ and} \\ |\Lambda_c(n) - \Lambda_c^*| &\leq 0.01, \end{aligned}$$

where  $\Lambda_h^* = \Lambda_w^* = \Lambda_c^* = 1/3$ , and  $m^*$  is the target slope (the target slope is similarly computed from the cumulative fatalities data in log scale; for example, the India fatalities' curve in the range

130–199 gives a slope of  $m^* = 0.1803$ ). Once the slopes are matched, assuming that the simulator starts on 01 March 2020, we find the delay between the fatalities curve from the simulator and the target data. We then use the resulting contact rates and the above calibration delay to launch our simulations.

To avoid any oscillatory behaviour of the calibration algorithm, we also set the scale factor in each of the above update steps to be  $[\exp((m^* - m(n))/n)]_a^{1/a}$  whenever  $|\Lambda_h(n) - \Lambda_h^*| \leq 0.02$ ,  $|\Lambda_w(n) - \Lambda_w^*| \leq 0.02$  and  $|\Lambda_c(n) - \Lambda_c^*| \leq 0.02$ . In addition, we set the scale factor to be  $[\exp((m^* - m(n))/(n - 25))]_a^{1/a}$  if  $n \geq 30$ .

For the simulation results presented in the case studies, we identified the  $m^*$  as follows. We assumed a *counterfactual* situation where the national level of infection, up to the date when the lockdown's effect is not yet likely to have been seen, estimated to be 8–10 April 2020, is moved to the city under study. There were 199 fatalities in India up to 10 April 2020. The India cases and fatalities data are based on daily updates compiled by the European Centre for Disease Prevention and Control<sup>27</sup>. The counterfactual situation (all of India's infections in the isolated city under study) is to ensure that sufficient data are available for calibration before the national lockdown's effect is encountered. The fatalities up to this date were likely due to contacts prior to the start of the lockdown. The slope of this curve (in the log domain) gives the  $m^*$ . The calibration is further done on a smaller version of the city, with 10 lakh population. The resulting parameters are then used on the full-scale city.

We do not calibrate  $\beta_T$ , the transmission coefficient at transport space. For the calibration step, we take this parameter to be zero while tuning the other parameters. A heuristic justification is as follows. Bengaluru travel interactions will likely be captured through the local community interactions, and we keep it zero throughout, even in the case studies. For Mumbai, however, local trains are a key mode of daily transportation with a population of the order of 75 lakh travelling daily using this mode in normal times. However, trains were stopped in Mumbai prior to the national lockdown and were running below capacity for at least a week before that. Moreover, the initial infections were seeded by travellers that came from abroad. The primary mode of travel for this group is unlikely to be rail transport. Therefore, we disabled the transport space while calibrating by setting  $\beta_T = 0$ . Subsequently, for the trains on/

off case study, we used a heuristic calculation of  $\beta_T$ ; see<sup>28</sup>, Section IV.

The above procedures identify the contact-related parameters. Other parameters are the distance kernel parameters, the parameter  $\alpha$  that accounts for crowding in households, the age-stratified interactions, the distribution parameters for individual infectiousness, the probability of severity, etc. These are set as follows (Table 7).

Among the above non-calibrated parameters, we have picked subnetwork upscale factor, project subnetwork size range, family friends' subnetwork range, and crowding factors somewhat arbitrarily. We need to conduct surveys to fix them to more realistic values. The parameters of the distance kernel  $f$  are obtained from fitting the origin–destination matrix to the model. The parameters related to the disease progression are obtained from the literature; see Sect. 4.2.

The compliance factors are chosen based on trial-and-error procedure as follows. Our current calibration methodology does not tune the compliance factor. Therefore, we pick a compliance factor, tune the model for transmission coefficients, run the simulator, and see how it matches with ground truth. We then repeat this procedure for various values of compliance factor, and then pick a reasonably good match by visual examination of the cases as well as the fatalities plots. While we pick the compliance factor each time, we use some heuristics to narrow down the search. For instance, in the early days of nation-wide lockdown in Bengaluru, one expects a higher compliance due to stricter enforcement of restrictions. Thus, we pick a higher compliance before 03 May 2020 and a lower compliance starting from 04 May 2020. In Mumbai, there is an additional distinction between households in the high-density areas and the low-density areas. To account for the fact that many households in the high-density areas may still need to access shared facilities, we choose a lower compliance for those households (once again, by just a visual examination of plots).

#### 4.6 Interventions

The simulator has the capability to accommodate interventions and compliance. Table 8 describes some of the interventions in<sup>6</sup>, some adapted to suit our demographics, and some new interventions involving the nation-wide 40-day 'lockdown' in India and various scenarios of 'unlock'. These are fairly straightforward to



implement—we modulate an individual’s contact rate with an interaction space (both into the interaction space and out of the interaction space) by a suitable factor associated with intervention. For example, one could easily implement and study cyclic exit strategies as done in<sup>29</sup>. The triggers for cyclic controls could be based on signals such as the number of individuals that are hospitalised, as done in our soft ward-containment. Yet, another one is to quarantine or case isolate based on contact tracing, as we will describe next (Table 8).

#### 4.7 Contact Tracing

Our simulator also includes a framework to study the impact of early contact tracing and testing. We assume that contacts of an individual in the smaller networks such as neighbourhood network, project network, class network, and random community network can be identified and tested/quarantined. The current contact tracing protocol quarantines certain primary contacts and tests a subset of these (e.g., symptomatic primary contacts). In our implementation, based on our study of ICMR’s testing protocol, given an index case, all household members, a fraction of the friends circle, a fraction of the inner school/workplace circle, and a fraction of the neighbourhood community are termed as primary contacts of this index case. All of these are quarantined, and a fraction of the symptomatic and another fraction of the asymptomatic among these are tested. Those who test positive become new index cases and spawn further contact tracing. The testing fractions are calibrated to match the actual reported cases and the test-positivity rate.

#### 4.8 Limitations

We list some limitations of our simulator.

- We do not have activity modelling in our simulator. As a consequence, weekly and daily patterns on interactions are not taken into account; for instance, the absence of interaction in workplaces and schools during weekends/public holidays, an increased interaction in public transport during morning and evening peak hours, etc. are not taken into account in our model. Instead, all these factors are abstracted into a single infection rate for each individual prescribed by (1), (3) and (4).
- Some of the data that we need in our simulator, such as the household-size distribution, workplace size distribution, school-size distribution,

commuter distance distribution, etc., can perhaps be difficult to obtain for some cities.

- We have too many free parameters in our model. This can lead to overfitting resulting in high generalisation error.
- The framework is computationally intensive.
- In our simulations, testing probabilities remain fixed throughout. We pick the testing probabilities, by trial-and-error, to match the number of tests and the test-positivity rate with the average number of tests and the average test-positivity rate, respectively, in Mumbai during the second week of July 2020. We then use these testing probabilities throughout the simulation. Therefore, if there is a significant change in the number of tests done compared to the situation during the second week of July 2020, our predictions in the distant future can be significantly different from the ground truth. A newer version of our simulator can be configured for the number of tests on each day and the test-positivity rate.

## 5 Algorithmic Aspects

### 5.1 Algorithmic Aspects Related to City Generation

Generation of a synthetic city is performed via the following steps.

1. *Data gathering and data preparation* involve the following.
  - (a) *Census Data Processing*: The primary data sources used to generate the synthetic city are the 2011 decennial census data of India and the intermediate survey reports for a city. The raw data are typically either in spreadsheets or as tables in a PDF document. We use Python packages like `pandas` and `tabula` to clean, process, and prepare the data required for creating a synthetic city.

The data required for creating a synthetic city like the ward-wise demographics, employment status, number of households, and the origin-destination matrix are created as separate comma separated values (csv) files. Distributions, such as the household size, workplace size, and school size, for

the city are collected as a single JavaScript Object Notation (JSON) file.

- (b) *Geo-spatial mapping*: In addition to the census data, the instantiation for the city also requires the geographic representations for a ward like the ward centre and ward boundaries. These are obtained from map files. Map files are mined in different formats like shapefiles (.shp, .shx) or geoJSON (.geojson), which are processed using Python's geopandas package.
2. *Instantiating the city files* is done by running a python script on the following inputs: the map data (.geojson), the census data on the demographics(.csv), households(.csv), and employment(.csv) files along with the additional parameters specified in 'cityProfile.json'. The instantiation of a synthetic city is done in three stages, namely:
- (a) *Processing Inputs*: The script ingests the input (.csv) files using the pandas package and computes parameters based on the input like the unemployment fraction, fraction of population in each ward. The GeoJSON file is processed with the geopandas package to parse the input file and the shapely package to compute the ward centre, the ward boundaries, and the neighbouring wards. Apart from the data files, the target population for which the instantiation is to be done, the average number of students in a school, the average number of individuals in one workplace are input parameters specified at the start of the script. The age distribution, household-size distribution, and school-size distribution are taken as inputs from the 'cityProfile.json' file.
  - (b) *Instantiating individuals* comprises of an algorithm that randomly assigns individuals to households by respecting the household-size distribution. Each household has individuals assigned with generational gaps, yet the instantiated population's empirical age distribution must match the given age distribution. Individuals are assigned to workplaces or schools or neither, based on their age and the unemployed fraction, and are assigned the appropriate 'workplaceType'. Once an individual is assigned to a household, the location of the individual is mapped to the

location of the assigned household. While instantiating households, the ward number to which the household is assigned is specified, and based on the ward number, the respective ward boundaries are obtained from the map data in the GeoJSON file. The ward boundaries are typically represented as 'Polygons'. The locations of households, workplaces, and schools are randomly sampled as a point inside the ward boundary. For instantiating households in high-density areas, we sample locations either from a GeoJSON file with boundaries of the high-density areas or from a collection of pre-sampled locations of households in high-density areas. Common areas where community interactions take place are instantiated at the ward centres, assumed to be the centroids of the polygons. These tasks are accomplished using the following python packages: numpy, random, pandas, and shapely. The outputs of this stage are collections of the instantiated individuals, their assigned households, schools, workplaces, transport, and community areas.

- (c) *Additional processing for generating city files*: Before generating the city files, additional processing is done on the dataframes which includes computing the distance of the individuals to their respective ward centres. This stage uses the pandas package for processing and generating the city files in the JSON file format for each instantiated collection namely the individuals, households, workplaces, schools, community centres, and distance between wards.

## 5.2 Algorithmic Aspects Related to Disease Spread

The disease progression part of the simulator is broadly implemented as follows. There are four time steps on a given day. At each time step, we go through each susceptible agent and find out the infection rate given by either (1), (3), or (4) depending on the interaction that we want to model. We then update the disease progression status of that agent based on the infection rate. Once an individual becomes exposed, this person goes through various stages of the disease depending on age, and contributes to the infection rate of other individuals in the individual's

interaction space as long as the individual is infectious. Eventually, the agent recovers from the disease or dies. At the end of the simulation, we output a time series of various quantities of interest into a file, such as the daily number of cases, cumulative number of cases, daily number of hospitalisations, daily number of fatalities, cumulative number of fatalities, etc.

Some of the key features of our simulator that help to reduce the space and time complexity are as follows.

- We have a single community space per ward where individuals living in that ward come together and interact. We then have an interaction among communities to model community interactions among people living in different wards. With  $n$  agents and  $c$  communities, such a model keeps the computational complexity at  $O(n) + O(c^2)$ . If we had considered a more complex interaction between individuals, where each individual interacts with every other individual living in the city with a certain contact rate, the complexity would have been  $O(n^2)$ . Modelling of other interaction spaces such as households, schools, workplaces, and transport spaces also results in a similar reduction in time complexity from  $O(n^2)$  to  $O(n)$ . Since the number of agents are typically of the order of  $10^7$ , such a reduction has a huge impact on the running time of the simulator.
- Contact tracing requires us to maintain a list of contacts made by each agent. In our implementation, we assume that each individual has a certain number of contacts that we can trace (which is random, but independent of  $n$ ). As a result, the space complexity becomes  $O(n)$  instead of  $O(n^2)$ .
- In the age-stratified interaction as well as OD-matrix-based distance kernel, we consider dominant terms of the age-based contact rate matrix as well as OD matrix by doing a principal component analysis and by focusing on a few important components. This helps to simplify the summations in (3).

These optimisation features appear to be novel features of our simulator.

## 6 Conclusion

In this work, we built an agent-based simulator to study the impact of various non-pharmaceutical interventions in the context of the ongoing

COVID-19 pandemic. We demonstrated the capabilities of our simulator via various case studies for Bengaluru and Mumbai. Some of the key features of our simulator include: (a) age-stratified interactions that capture heterogeneity in interaction among people in a given interaction space; (b) the ability to implement various interventions such as soft ward-containment, phased opening of workplaces, and community spaces; (c) a broad class of contact tracing-based testing and case isolation protocols, etc. These features allow our simulator to capture many aspects of the ground reality which are important for comparative analysis and realistic predictions.

However, as with any exercise in modelling a complex system, our current approach has several limitations. Some of these limitations were discussed in detail in Sect. 4.8. We now suggest two broad directions for future work. The first of these relates to enhanced data collection to identify with greater accuracy important parameters such as disease progression statistics, rates of compliance to various interventions, the effect of public transit systems, etc. The second relates to enhancing the model itself in various ways so as to make it a more versatile tool for administrators and epidemiologists. Some of these possible enhancements may include bringing in movement of people into and out of the city and studying the impact of various mobility patterns, modelling and studying the impact on public-health oriented decisions on the economy, incorporating more fine-grained activity modelling, and, finally, extending the simulator to obtain district-scale or country-scale predictions. We hope that such work will lead to agent-based simulators finding a regular place in every public-health official's tool kit.

## Publisher's Note

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

## Acknowledgements

We sincerely thank Arpit Agarwal, Priyanka Agrawal, Bharadwaj Amrutur, M. S. Ashwin, Siva Athreya, Abhijit Awadhya, Chiranjib Bhattacharyya, Vijay Chandru, Avhishek Chatterjee, Arpan Chattopadhyay, Siddhartha Gadgil, Aditya Gopalan, K. V. S. Hari, Ramesh Hariharan, Aniruddha Iyer, Srikanth Iyer, Shipra Jain, Subhasis Jethy, Jacob John, Shruti Kakade, Navin Kashyap, Micky Kedia, DataMeet (a community of Data Science and Open Data enthusiasts), Revanth Krishna,

Vrishab Krishna, Manjunath Krishnapur, Anurag Kumar, Rahul Lodhe, Rahul Madhavan, Madhav Marathe, Satyajit Mayor, Gautam Menon, Shiraz Minwalla, Prem Mony, Chandra Murthy, Y. Narahari, Uma Chandra Mouli Natchu, Praneeth Netrapalli, Abdul Pinjari, Vinod Prabhakaran, Dr. Prabhu, Rishi Prajapati, Ketan Rajawat, G. Rangarajan, G. V. Sagar, Agrim Sharma, Abhishek Sinha, Mukund Thattai, Himanshu Tyagi, Srinivasan Venkataramanan, V. Vinay, Anil Vullikanti, and many volunteers for their valuable comments and help.

The SahasraT supercomputing cluster at the Supercomputer Education and Research Centre (SERC), Indian Institute of Science, was used for most of the simulations.

PP, NR, SS, NKV, and SY from IISc were supported by the IISc-Cisco Centre for Networked Intelligence, Indian Institute of Science. RSun was supported by the IISc-Cisco Centre for Networked Intelligence, the Robert Bosch Centre for Cyber-Physical Systems, and the Department of Electrical Communication Engineering, Indian Institute of Science. TIFR co-authors acknowledge support of the Department of Atomic Energy, Government of India, under project no. RTI4001. PH acknowledges support from the Swarnajayanti Fellowship of DST. RSap acknowledges support from the Ramanujan Fellowship of SERB. PS acknowledges support from the Ramanujan Fellowship of SERB, from Adobe Systems Inc. via a gift to TIFR, and from the Infosys-Chandrasekharan virtual centre for Random Geometry supported by a grant from the Infosys Foundation. The contents of this paper do not necessarily reflect the views of the funding agencies.

Received: 18 October 2020 Accepted: 19 October 2020  
Published online: 12 November 2020

## References

1. Ministry of Health and Family Welfare, Government of India. [Online]. <https://www.mohfw.gov.in/>. Accessed 11 Aug 2020
2. Kermack WO, McKendrick AG, (1927) A contribution to the mathematical theory of epidemics. In: Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, vol 115(772), pp 700–721
3. Shekatkar S, Pujari B, Arjunwadkar M, Hazra DK, Chaudhuri P, Sinha S, Menon GI, Sharma A, Guttal V (2020) INDSICI-SIM A state-level epidemiological model for India, ongoing Study at <https://indscicov.in/indscisim>
4. Prakash MK, Kaushal S, Bhattacharya S, Chandran A, Kumar A, Ansumali S (2020) A minimal and adaptive prediction strategy for critical resource planning in a pandemic. medRxiv
5. Ganesan S, Subramani D (2020) Spatio-temporal predictive modeling framework for infectious disease spread, arXiv preprint [arXiv:2006.15336](https://arxiv.org/abs/2006.15336)
6. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunuba Perez Z, Cuomo-Dannenburg G et al (2020) Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand Tech. Report
7. Gardner JM, Willem L, van der Wijngaart W, Kamerlin SCL, Brusselaers N, Kasson P (2020) Intervention strategies against covid-19 and their estimated impact on swedish healthcare capacity. medRxiv
8. Halloran ME, Ferguson NM, Eubank S, Longini IM, Cummings DA, Lewis B, Xu S, Fraser C, Vullikanti A, Germann TC et al (2008) Modeling targeted layered containment of an influenza pandemic in the United States. Proc Natl Acad Sci 105(12):4639–4644
9. Hunter E, Mac Namee B, Kelleher JD (2017) A taxonomy for agent-based models in human infectious disease epidemiology. J Artif Soc Soc Simul 20:3
10. Eubank S, Guclu H, Kumar VA, Marathe MV, Srinivasan A, Toroczkai Z, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. Nature 429(6988):180–184
11. Athreya S, Gadhiwala N, Mishra A (07 2020) COVID-19 India-Timeline: An Understanding across States and Union Territories, [Online]. Available: <https://www.isibang.ac.in/~athreya/incovid19/>
12. Crowdsourced COVID19-India Database, <https://www.covid19india.org/>. Accessed 04 2020
13. MacIntyre CR, Cauchemez S, Dwyer DE, Seale H, Cheung P, Browne G, Fasher M, Wood J, Gao Z, Booy R, Ferguson N (2009) Face mask use and control of respiratory virus transmission in households. Emerg Infect Dis 15(2):233–241. <https://doi.org/10.3201/eid1502.081167>
14. Aiello AE, Murray GE, Perez V, Coulborn RM, Davis BM, Uddin M, Shay DK, Waterman SH, Monto AS (2010) Mask use, hand hygiene, and seasonal influenza-like illness among young adults: a randomized intervention trial. J Infect Dis 201(4):491–498. <https://doi.org/10.1086/650396>
15. Leung N, Chu D, Eea Shiu (2020) Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat Med 26:676–680. <https://doi.org/10.1038/s41591-020-0843-2>
16. Chu D, Akl E, El-Harakeh A, Bognanni A, Lotf T, Loeb M, Hajizadeh A, Bak A, Izcovich A, Cuello-Garcia CA et al (2020) Physical distancing, face masks, and eye

- protection to prevent person-person covid-19 transmission: A systematic review and meta-analysis. *Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9)
17. Consolidated Revised Guidelines on the measures to be taken by Ministries/Departments of Government of India, State/ UT Governments and State / UT authorities for containment of COVID-19 in the country, 15 Apr 2020. [Online]. [https://www.mha.gov.in/sites/default/files/MHA%20order%20dt%2015.04.2020%2C%20with%20Revised%20Consolidated%20Guidelines\\_compressed%20%283%29.pdf](https://www.mha.gov.in/sites/default/files/MHA%20order%20dt%2015.04.2020%2C%20with%20Revised%20Consolidated%20Guidelines_compressed%20%283%29.pdf)
  18. MHA order dt. 30.5.2020 on phased re-opening (unlock 1), 30 May 2020. [Online]. [https://www.mha.gov.in/sites/default/files/MHAOrderDt\\_30052020.pdf](https://www.mha.gov.in/sites/default/files/MHAOrderDt_30052020.pdf)
  19. COVID-19 Community Mobility Report. <https://www.google.com/covid19/mobility/>. Accessed 04 2020
  20. Chandak A, Dey D, Mukhoty B, Kar P (2020) Epidemiologically and socio-economically optimal policies via Bayesian optimization . [Online]. <https://arxiv.org/abs/2005.11257>
  21. Baker J, Basu R, Cropper M, Lall S, Takeuchi A, (2005) *Urban poverty and transport : the case of Mumbai*. The World Bank [Online]. <https://doi.org/10.1596/1813-9450-3693>
  22. Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsrithaworn S, Burke DS (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437(7056):209–214
  23. Prem K, Cook AR, Jit M (2017) Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol* 13(9):e1005697
  24. Singh R, Adhikari R (2020) Age-structured impact of social distancing on the COVID-19 epidemic in India, arXiv preprint [arXiv:2003.12055](https://arxiv.org/abs/2003.12055)
  25. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PG, Fu H et al (2020) Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*
  26. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CA, Munday JD et al (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*
  27. European Centre for Disease Prevention and Control's COVID-19 website. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>. Accessed 04 2020
  28. IISc-TIFR Covid-19 City-Scale Simulation Team, COVID-19 Epidemic Study II: Phased Emergence From the Lockdown in Mumbai," Indian Institute of Science and Tata Institute of Fundamental Research, Tech. Rep., June 2020. [Online]. <https://arxiv.org/pdf/2006.03375.pdf>
  29. Karin O, Bar-On YM, Milo T, Katzir I, Mayo A, Korem Y, Dudovich B, Yashiv E, Zehavi AJ, Davidovich N et al, Adaptive cyclic exit strategies from lockdown to suppress COVID-19 and allow economic activity. *medRxiv* (2020)



**Shubhada Agrawal** is a Ph.D. student at the School of Technology and Computer Science at TIFR, Mumbai. Her research interests are in applied probability, learning theory, and Monte Carlo methods for simulations.



**Siddharth Bhandari** is a graduate student at STCS, TIFR Mumbai. His research interests lie in the area of Randomised Algorithms.



**Anirban Bhattacharjee** is a student at the School of Technology and Computer Science at TIFR Mumbai. His research interests are in applied probability, reinforcement learning and covid science.



**Anand Deo** is a PhD student at the School of Technology and Computer Science, Tata Institute of Fundamental Research. His research interests are in Applied Probability, Operations Research, Machine Learning, and Stochastic Modelling.



**Narendra M Dixit** is Chair of the Centre for Biosystems Science and Engineering and Professor of Chemical Engineering at the Indian Institute of Science, Bengaluru. His research interests are in computational biology of infectious diseases including HIV and COVID-19.



**Prahladh Harsha** is a theoretical computer scientist at the Tata Institute of Fundamental Research (TIFR). He received his B.Tech. degree in Computer Science and Engineering from the IIT Madras in 1998 and his S.M. and Ph.D. degrees in Computer Science from MIT in 2000 and 2004 respectively. Prior to joining TIFR in 2010, he was at Microsoft Research, Silicon Valley and at the Toyota Technological Institute at Chicago. His areas of interests include computational complexity, hardness of approximation, coding theory and information theory.



**Sandeep Juneja** is a Professor and Dean at the School of Technology and Computer Science in Tata Institute of Fundamental Research in Mumbai. He received his B.



Tech. in Mechanical Engineering from IIT Delhi (1989) and his M. S. in Statistics and Ph.D. in Operations Research from Stanford University (1993). He has been with TIFR since 2003. He is currently on the editorial board of Stochastic Systems. Earlier he has been on editorial boards of Mathematics of Operations Research, Management Science and ACM TOMACS. His research interests lie in applied probability including in sequential learning, mathematical finance, Monte Carlo methods, and game theoretic analysis of queues. Lately, he has been exploring some aspects of modelling epidemic spread.



**Poonam Kesarwani** is a visiting fellow at the STCS Department at TIFR, Mumbai. Her research interests are in applied probability, machine learning, optimization theory.



**Aditya K Swamy** worked as a HPC specialist at the Supercomputer Education and Research Centre (SERC), IISc, training users in HPC and collaborating with them for performance optimization of applications. His other research interests are in Computational Physics and Plasma Physics, in which he has a PhD from the Institute for Plasma Research, Gandhinagar. Since August, he works for AMD India.



**Preetam Patil** is a project staff at the Centre for Networked Intelligence [RBCCPS-EECS division]. His research interests are in (software-defined/wireless/data-center) computer networks.



**Nihesh Rathod** is a Cisco Research Scholar in Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, where he is currently pursuing his Ph.D. degree in communication and networks. His interest include the areas in design and implementation of Internet of Things (IoT).



**Ramprasad Saptharishi** is a Reader at the School of Technology and Computer Science at the Tata Institute of Fundamental Research, Mumbai. His research interests are primarily complexity theory, coding theory and algebraic methods.



**Sharad Shriram** is a staff at the Centre for Networked Intelligence at the Indian Institute of Science, Bengaluru. His research interests broadly lie in Human-centered AI, Natural Language Processing, Edge Intelligence, and Data-driven Modelling.



**Piyush Srivastava** is a Reader at the School of Technology of Computer Science at the Tata Institute of Fundamental Research, Mumbai. His research interests are in Theoretical Computer Science.



**Rajesh Sundaresan** received the B.Tech. degree in electronics and communication from IIT Madras, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, in 1996 and 1999, respectively. From 1999 to 2005, he worked on the design of communication algorithms for wireless modems with Qualcomm Inc. Since 2005, he has been with the Indian Institute of Science, where he is currently a Professor with the Department of Electrical Communication Engineering, an Associate Faculty Member of the Robert Bosch Centre for Cyber-Physical Systems, and a member of the Centre for Networked Intelligence. His research interests include the areas of communication, computation, and control over networks, applied probability and statistics.



**Nidhin Koshy Vaidhiyan** is a consultant technologist at the Centre for Networked Intelligence at the Indian Institute of Science, Bengaluru. His research interests are in wireless communications and communication networks.



**Sarath Yasodharan** is a PhD student at the Department of Electrical Communication Engineering at the Indian Institute of Science, Bengaluru. His research interests are in applied probability.