# A Novel Systems Model to Simulate the Dynamics for the Analytical Study of the Proliferation of COVID-19

**Arpita Welling**
Indiana University, Bloomington, Indiana, USA
**Abhilasha Patel**
Pi Innovate, Applied Research Group, Pune, Maharashtra, India
**Padmaj Kulkarni**
Deenanath Mangeshkar Hospital, Dept. of Medical Oncology, Pune, Maharashtra, India
**Vinay G. Vaidya** ( vinay.vaidya@piinnovate.com )
Pi Innovate, Applied Research Group, Pune, Maharashtra, India https://orcid.org/0000-0002-6562-6261

**Research Article**

**Keywords:** MIMANSA, Systems Modeling, Mathematical modeling, COVID-19, SARS-CoV-2, Complex biological models

**Posted Date:** October 14th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-968078/v1

# A Novel Systems Model to Simulate the Dynamics for the Analytical Study of the Proliferation of COVID-19

**Arpita A. Welling**[1], **Abhilasha P. Patel**[2], **Padmaj S. Kulkarni**[3], **and Vinay G. Vaidya**[2, 4, *]

[1]Indiana University, Bloomington, Indiana, USA
[2]Pi Innovate, Applied Research Group, Pune, Maharashtra, India
[3]Deenanath Mangeshkar Hospital, Dept. of Medical Oncology, Pune, Maharashtra, India
[4]Dept. of Technology, Savitribai Phule Pune University, Pune, Maharashtra, India
[*]vinay.vaidya@piinnovate.com

## ABSTRACT

In this paper, we present a systems model for COVID-19 called 'Multilevel Integrated Model with a Novel Systems Approach' or MIMANSA for short. This model goes to the individual patient level and mimics small steps in the process of virus spread. Despite many models simulating the growth of COVID-19 cases, it is rare to see a comprehensive model that takes into account the transmissibility of variants, the prevalence of multiple variants, growth of the variants, percentage of vaccinated population, exposure rate, infection rate, silent carrier rate, secondary infection rate, mask usage, and people mobility all in one model.

We begin by categorizing the in-person social interactions of an individual into three areas: household, workplace, and public places. With each interaction, the virus spreads from an infectious person to a healthy individual. We build the model, one level at a time, covering daily person-to-person interactions. Further, MIMANSA forms a new layer of network for every new day. These layers form a part of the virus proliferation network.

MIMANSA models the virus incubation period using a Weibull distribution. Once the network starts building, the virus growth can be curtailed only through increased vaccinations or through non-pharmaceutical interventions such as mask usage, reduced mobility, and/or quarantine.

MIMANSA takes the mobility data, mask usage data, variant prevalence data, and vaccination data as inputs. Despite the model being intricate, it uses only 5 parameters for training. Once the model is trained, it can be slightly adjusted by the daily environmental variable. This variable corrects for day to events as well as varying environmental conditions in a given location.

We present the results of the training and validation for the USA, California, and the UK. It is seen that during the validation stage, the model accuracy is within 2%. Further, projections are made for about 3 weeks. In the end, we have presented the results of a study on the effect of vaccination on the number of COVID cases in the USA. It shows how MIMANSA can be used for studying the impact of multiple scenarios. Additionally, the model is not only useful for making predictions in the number of cases but also useful as an educational aid for explaining the proliferation network of the virus in real life. Although MIMANSA is originally developed for the SARS-CoV-2, it can be modified to study the spread of any other virus, and in any region.

## Introduction

A good growth model increases one's ability to understand the underlying phenomenon that makes the virus growth possible. Growth projections should never be the sole objective of a model but it should be an obvious outcome or a byproduct of a good model. Understanding the biological phenomenon, formulating it in terms of mathematical equations that are logically interlinked to mimic the observed facts, forms a good foundation of any good biological model. Mathematical modeling of COVID-19 (Coronavirus Disease 2019) is no exception.

Over the past years, researchers all over the world have increased our understanding of how the SARS-CoV-2 grows. This includes some basic understanding such as the impact of social distancing, mask usage, and patient isolation. Further, our knowledge advanced to the exposure rate, silent carrier rate, infection rate, virus transmission rate, secondary attack rate, and variant growth. Then came the vaccines and we learned how the cases go down with the increase in vaccinated people. Thus, one can build a mathematical model that makes use of this information. Additionally, the model should be universally

applicable with due considerations to regional exposure variation factors, and a small daily variability in exposure accounting for local events.

There are many attempts made to develop mathematical models of COVID-19, however, they do not always answer questions related to the impact of multiple variants, vaccination, mask usage, and government policies. In this paper, we present a novel approach for mimicking the SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) infection by interconnecting multiple mini models to form one systems model. Our model, Multilevel Integrated Model with a Novel Systems Approach, MIMANSA, is built from epidemiological observations and data. It is well known that studying the asymptomatic population is vital in the study of the spread of COVID-19. The simulation of the virus spread through silent carriers is at the heart of MIMANSA. The MIMANSA model can delineate between patients infected with the virus, pre-symptomatic infectious patients, symptomatic patients as well as those that are asymptomatic. Further, MIMANSA distinguishes the degree of spread based on an individual's interaction between three different groups, namely individuals at home, workplace, and public places. It considers the percentage of vaccinated people, prevalence of multiple variants, mask usage, and mobility data to train the model.

It has provisions for simulating non-pharmaceutical interventions such as a lockdown, mask usage, and household quarantining. It can simulate scenarios to study the impact of precautions such as mask usage. It includes the transmissibility of new variants of SARS-CoV-2 as well as the impact of vaccination.

## Literature Survey

There are several papers in the literature that use the mathematical modeling approaches to forecasting COVID-19 cases. This survey gives a quick overview of the broad areas in mathematics applied towards simulating the COVID-19 spread.

There is a misconception that the COVID-19 cases grow exponentially. Although the exponential model is one of the most widely used models for the growth of any organism, the growth in COVID-19 patients does not follow the exponential curve throughout. Any rapid growth may appear to be exponential growth. However, it is only so in the beginning. The fallacy in presuming the growth of COVID-19 cases to be exponential becomes apparent if one takes a look at the data for a period of 6 to 8 months.

Medina-Mendieta et al.[1] use Gompertz, and Logistic equations for modeling the growth of COVID-19; however, these equations do not help in considering multiple scenarios that arise in the management of COVID-19 such as the lockdown, quarantine, exposure variation, etc. The main objective in building a good model is in getting insights that one cannot get otherwise. The objective is not merely projecting the number of cases by using an equation.

Time series analysis, using the Auto Regressive Moving Average (ARMA) models, has been another popular approach. Among many researchers using the ARMA model, Maleki, Mohsen, et al.[2] have used ARMA for forecasting, and Malki, Zohair, et al.[3]. have used it to predict the end of COVID-19. The ARMA models are used for stochastic data. The number of COVID cases has a cause-and-effect relationship which is mostly deterministic with some stochastic uncertainty. Thus, it is a wrong choice of the model for predicting the growth of COVID cases.

Machine learning methods are useful in many applications. Khakharia, Aman, et al.[4] use machine learning to predict the outbreak of COVID-19 in highly populated and dense areas. There are many other approaches to forecast the spread of COVID-19. Kucharski et al.[5] studied the early dynamics of transmission. They modeled SARS-CoV-2 using a geometric random walk process and Monte Carlo simulation. Grassly and Fraser[6] have given a good review of mathematical models of infectious disease transmission. Klinkenberg et al.[7] developed a model for contact tracing. A large majority of the papers use the conventional SIR, SEIR, SIRD type of model with multiple variations. Mandal et al.[8] used the classical SEIR (Susceptible, Exposed, Infectious, and Recovered), model. Fanelli and Piazza[9] forecasted cases in China, Italy, and France based on the variation of the SEIR to the SIRD model. In this approach, the Susceptible (S), Infected (I), Recovered (R), and the Dead (D) model, every person who is going to be infected by the virus, falls in one of the four categories. Then the equations are set up as first-order differential equations with one differential equation per stage.

However, there are some drawbacks to the SIR group of models. These inadequacies are studied by Moein et.al.[10]. Their conclusion of the study was that the SIR model is reasonable in forecasting in the short run, however, it is inadequate to do so in the long run. The SIR model is not appropriate for simulating different features of the COVID-19. Thus, the COVID-19 needs a more sophisticated model that considers epidemiological and biomedical observations.

An ideal model should mimic the actual phenomenon by considering the facts available at the time of the development. It should also have the flexibility to add on and integrate any newly discovered phenomenon. The model should reveal something that one has not discovered yet. For these reasons, we gathered epidemiological facts related to the growth of SARS-CoV-2. Our model uses the systems approach to model development.

The systems approach was first suggested by Von Bertalanffy[11]. In this approach, a system is broken down into small subsystems. The subsystems are open to interacting with the environment as well as with each other. It is like watching a tree as well as being able to see the forest. One time you may see parts of the whole system while other times, you may like to

view the whole system from parts. Each subsystem is built on observed facts. All subsystems integrate and keep the observed phenomenon unperturbed. The systems approach has been effectively used for developing a model for tumor growth and chemotherapy[12]. Later, in 1991, the effect of fasting on reducing tumor growth by using a systems approach was studied[13]. Similarly, the model presented here, MIMANSA, is built using the system's principles.

## Mathematical Model

The MIMANSA model developed here is based on empirical data, clinical data from research papers, and global epidemiological observations of the COVID-19 cases. The observations used while building the model to explain the primary virus spread are given below.

- The virus spread occurs due to coming in contact with a virus-infected or virus-carrying person.
- Pre-symptomatic, asymptomatic, and healthy carriers are responsible for the spread of the virus. Asymptomatic patients can transmit the virus, even if they continue to show negative on the SARS-CoV-2 test[14–16].
- It is presumed that the symptomatic patients are isolated from the rest of the healthy people.
- Although the possibility is low, a healthy person may get infected if the person touches a fomite.
- For all reporting, the unit of time is one day. The total number of patients is reported every day. Similarly, new silent carriers are created every day.
- Since there is a 2 to 14-day incubation period for the SARS-CoV-2 virus, virus-infected persons do not show symptoms on day 1. It is observed that most patients start showing symptoms on the $5^{th}$, $6^{th}$, or $7^{th}$ day after being infected, with 97.5% showing symptoms by 11.5 days[17,18].

The following assumptions are made for the development of the model.

1. Infected patients spread the virus only during the infectious stage. This includes pre-symptomatic virus patients, asymptomatic patients after the incubation period, as well as healthy carriers with fomite.
2. When symptoms show, it is presumed that the patients will be hospitalized or effectively isolated/quarantined, eliminating them as a vector that can spread the virus to other healthy individuals.
3. The SARS-CoV-2 virus remains in aerosols for up to 3 hours, copper surfaces for up to 4 hours, and on cardboard for 24 hours. On plastic and stainless steel, it can remain for up to 2 to 3 days[19]. For simplicity, we consider the period of the virus remaining on a surface to be 24 hours.
4. To understand the spread of diseases, Del Valle et al.[20] studied how people from different age groups interact socially. In their study, they found that on average, between the ages of 20 and 50, a person meets 22 people every day. We have used this number to simulate the spread of SARS-CoV-2.
5. For a Variant of Concern (VOC) to be included in our study, its prevalence has to be greater than 10%. All the variants and their sub-lineages are assumed to have the same transmissibility. For instance, B.1.617.2 and AY.4 have the same transmissibility.
6. It is observed that masks are not 100% effective and there is always room for transmission of the virus. Moreover, not everyone wears an N95 mask with 95% efficacy[21]. Additionally, improper usage of a mask also reduces its effectiveness. To account for all these factors, we have assumed the mask effectiveness as 70%. Thus, we reduced the observed mask usage from the web-based data by 30%.

### Preliminaries

The terminology and naming convention used is an important step in understanding a mathematical model. In this section, we explain some basic terminology and relate it to the real-life scenario. The detailed list of the terminologies used in this study is explained in Table 1.

Virus Patients (VP) are the individuals marked in the simulation like the ones who would show up with symptoms during the incubation period. VPs can be of three types. VPs that are initially pre-symptomatic but not infectious, pre-symptomatic and infectious, and symptomatic. During the initial phase, VPs may not infect others. VPs are infectious virus patients from about 0.7 to 2.3 days before showing symptoms[22].

The pre-symptomatic patients who are infectious are labeled in MIMANSA as Unidentified Virus Patients or UVPs. UVPs are identified as virus patients when they become symptomatic. Symptomatic individuals are counted in the Daily Number of Cases (DNC) of COVID-19.

A silent carrier (SC) is an asymptomatic person who may never show symptoms. Although the silent carriers may not show symptoms, they can be infectious while being in the incubation period.

A Healthy Carrier (HC) is a person who is not infected with the virus but carries the virus via fomites. Fomites are materials such as a book, a bag, clothes, utensils, etc. that have been contaminated by the virus and serve as a mode of transmission. Remaining Healthy People (RHP) are the number of individuals who remain healthy despite being in contact with infectious individuals by the end of the day.

To incorporate the impact of vaccination on the COVID-19 growth, we consider the number of vaccinated people, variants present, and virus transmission rate. Accordingly, exposure-related parameters are adjusted.

### The Spread of the Infection

Individual interactions can broadly be divided into three categories, namely household, workplace, and public places. Consider that a group of households meet an external silent carrier E0 as shown in Fig. 1. During this meeting, P1 gets infected with the virus and is marked internally as a VP, while P2 becomes a silent carrier or asymptomatic patient and is marked as $SC_1$. P3 carries infected materials thus P3 is a healthy carrier, $HC_1$. P4 remains healthy and is counted as a remaining healthy person in the group. E0 interacts in three groups, namely household, workplace, and public places. It is important to note that there is only one virus-infected individual, i.e., E0 in the example, who interacts with people in all three groups. In all, E0 meets 22 people. All events are assumed to have taken place in one day, the time unit for the simulation cycle. This is how the virus spreads begins.

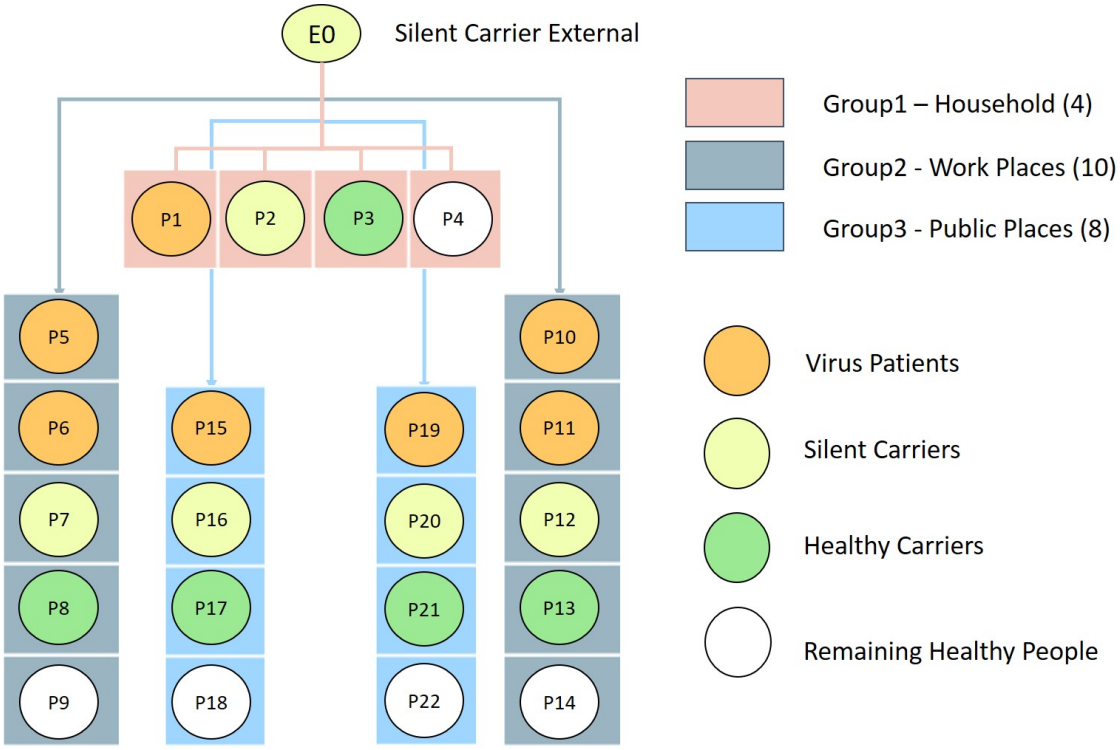



**Figure 1** Start of the virus spread through people interaction

In MIMANSA, at the micro-level, we have a block called the Internal Processing Diagram (IPD). IPD carries out the separation of patients in various categories and does some calculations as shown in Fig. 2. IPD forms a building block for the next level. At the next level, we have the Computational Block (CB) as in Fig. 3. The CB has multiple IPDs. The IPDs are used to build three groups namely household, workplace, and public places. These three groups form a computational block. Multiple CB blocks form a Virus Proliferation Network (VPN) as shown in Fig. 4. The VPNs have various levels. When a day is over, a new level starts with new CB blocks as shown in Fig. 4. At the end of the day, all symptomatic patients are counted and labeled as the daily new cases (DNC). At the end of the day, it is presumed that all symptomatic patients are either isolated in their home or a hospital. As soon as one day is over, the new day starts automatically and the network grows internally.

| Acronym | Explanation |
|---|---|
| CB | Computational Block |
| CB (with subscripts) | For example, CB23 is the computational block at level 2 and time 3. |
| CTQ | Contact Tracing and Quarantining used for household quarantining. Range 0 to 1. |
| DEV | Daily Environmental Variability Factor |
| DNC | Daily New cases |
| ER | Exposure Rate. Probability of getting exposed while interacting with silent carriers. The ER1, ER2, and ER3 are exposures for 3 groups. |
| G | In-person, social interaction Group |
| HB | Horizontal Block |
| HC | Healthy Carrier, a healthy person carrying a fomite. $HC_1$ are the Healthy Carriers, carrying fomites, coming from group 1. |
| HCR | Healthy Carrier Rate |
| IPD | Inter Processing Diagram for patient categorization and calculations |
| IPM | Infectious Patient Module |
| IR | Infection Rate is the probability of getting infected if one is exposed to the virus. |
| LKD | Values LKD1, LKD2, LKD3 are lockdown values for group1 (home), group2 (workplace), and group3 (public places). It is the fraction of the people following the lockdown orders. Range 0 to 1. |
| MUR | Mask Usage Ratio. Percentage of people using a mask. Ranges from 0 to 1. |
| N1, N2, N3 | The number of people in group1, group2, and group3 respectively. |
| NMUR | Non-Mask Usage Ratio. The fraction of people not using masks. It is (1-MUR). |
| P | P1, P2, P3, ....P14 are patients coming out of a PD |
| PD | Probability Distribution |
| PI | People Infected with the SARS-CoV-2 virus |
| $PQRHP_1$ | Post Quarantine Remaining Healthy People from group1 |
| PSTSC | Post PD Sum Total of Silent Carriers |
| REV | Regional Environmental Variability Factor, a suffix X is often added to REV i.e., REV_$X$ where X represents exposure rate, household quarantining, etc |
| RFSC | Secondary Attack Rate, Reduction Factor for Silent Carriers |
| RFUVP | Secondary Attack Rate, Reduction Factor for Virus Patients |
| RHP | Remaining Healthy People |
| RSUVP | Reduced SUVP |
| SC | $SC_1$, $SC_2$, $SC_3$ are the Silent Carriers for group1, 2, and 3 respectively |
| SCE | Silent Carriers External |
| SCR | Silent Carrier conversion Rate |
| SIP | Sum of Infectious Patients |
| SPTSC | Sum of last 14 levels |
| SSC | Sum of Silent Carriers |
| STSC | Reduced Sum of SSCs of last 14 levels at time t |
| SUVP | Sum of all Unidentified Virus Patients in the last 14 days. |
| SVP | Sum of Virus Patients |
| SYMVP | Symptomatic VPs |
| t, T | t is the time, in days, from the start. T is the incubation period time |
| TCTQ | Time at which Contact Tracing and Quarantining is enforced. |
| UVP | Unidentified Virus Patients. They are pre-symptomatic patient |
| VB | Vertical Block |
| VOC | Variant of Concern |
| VP | $VP_1$, $VP_2$, $VP_3$ are the Virus-infected Patients for group1, 2, and 3 respectively |
| VPPD | Virus-infected Patients who would come out of the PD showing symptoms |

**Table 1** List of abbreviations

## Methods and Materials

In the following sections, we give details of each building block. Further, we explain how one block is used to create the next bigger block. This bigger unit is further expanded into a network. Similarly, the virus spreads its network.

### Internal Processing Diagram (IPD)

Fig. 2 shows an Internal Processing Diagram (IPD). This diagram depicts what happens to every healthy person who comes in contact with an infectious person during group interactions. Inside the Internal Processing Diagram, every healthy person is exposed to the SARS-CoV-2 virus. Anyone coming inside the IPD comes out as either a Virus-infected Patient (VP), a Silent Carrier (SC), or a Healthy Carriers (HC). Every IPD symbol has these three outputs. Some people continue to remain healthy despite getting exposed to a silent carrier, and they are marked as the Remaining Healthy People (RHP).

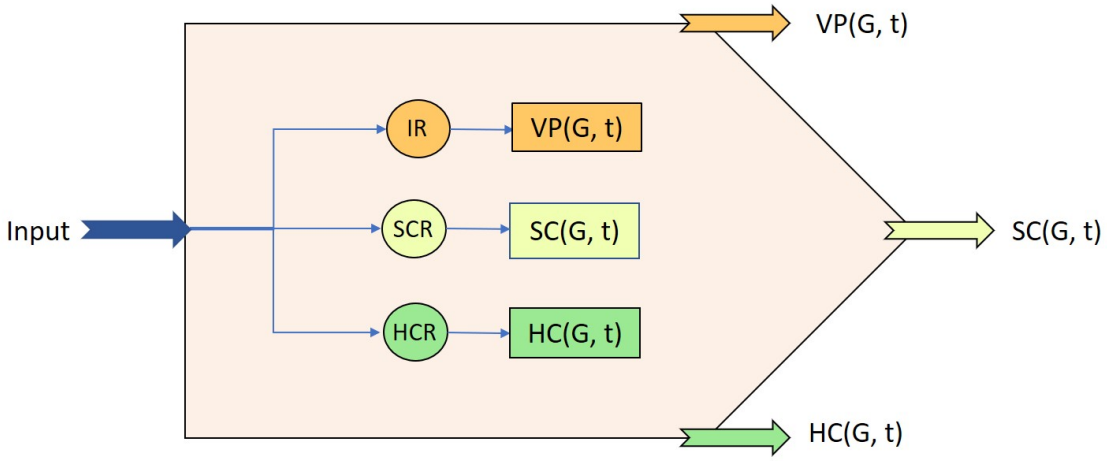



**Figure 2** Internal Processing Diagram (IPD)

IR is the infection rate that indicates the probability of becoming a virus patient if one is exposed to the virus. The SCR is the probability of conversion of a healthy individual into a silent carrier or an asymptomatic patient. The HCR is the rate of conversion of a healthy person into a healthy carrier.

### Computational Block

Fig. 3 shows the internals of a Computational Block (CB) which represents an individual's interaction with people in three different groups like household (group1), workplaces (group2), and public places (group3). These groups are denoted by three horizontal blocks stacked sequentially. When a person goes to these three places s/he meets an average of N1 people in a group1, N2 people in a group2, and N3 people in a group3.

The vertical block carries out some logical checks to select input to different groups in the CB. Inputs to the three groups come either from the total Patients who are Infectious (PI) or from the number of people who remained healthy on the previous day, RHPs.

PI is the count of infectious people. These are the people who meet different people the next day. Therefore, PIs are multiplied by the number of people they meet in each group. RHPs, the remaining healthy people, are labeled as per their group of origin 1, 2, or 3. RHPs are the people who are not yet infected but are in contact with the same infectious people and therefore they are not multiplied by the number of people in the group. RHPs are the balance number of people from the previous day. If quarantining policy is implemented, when any virus patient is detected, all household members would be quarantined. This reduces the number of RHPs on the next day. Thus, $RHP_1$s are reduced by a factor to give the count of Post Quarantine Remaining Healthy People (PQRHP). It is further explained in the section titled Quarantining.

The exposure rate is the probability of getting exposed to the virus while interacting with infectious individuals. For the three groups, namely home, workplace, and public places, we define three types of Exposure Rates, ER1, ER2, and ER3 respectively.

Lockdowns, also known as stay-home orders, are location-dependent. We label lockdown in the household, workplace, and public places as LKD1, LKD2, and LKD3 respectively. The LKD values are from 0 to 1. Thus, during a lockdown period, the fraction of people at workplaces is given by (1-LKD2) and in public places, it is (1-LKD3).

The healthy carriers from group 1, $HC_1$, also go to workplaces and public places. Thus, they are given as inputs to group 2 and group 3. All processing takes place in IPD, the inter-processing diagram. The output of the CB block is the number of VPs, SVP, SSC, and RHP.

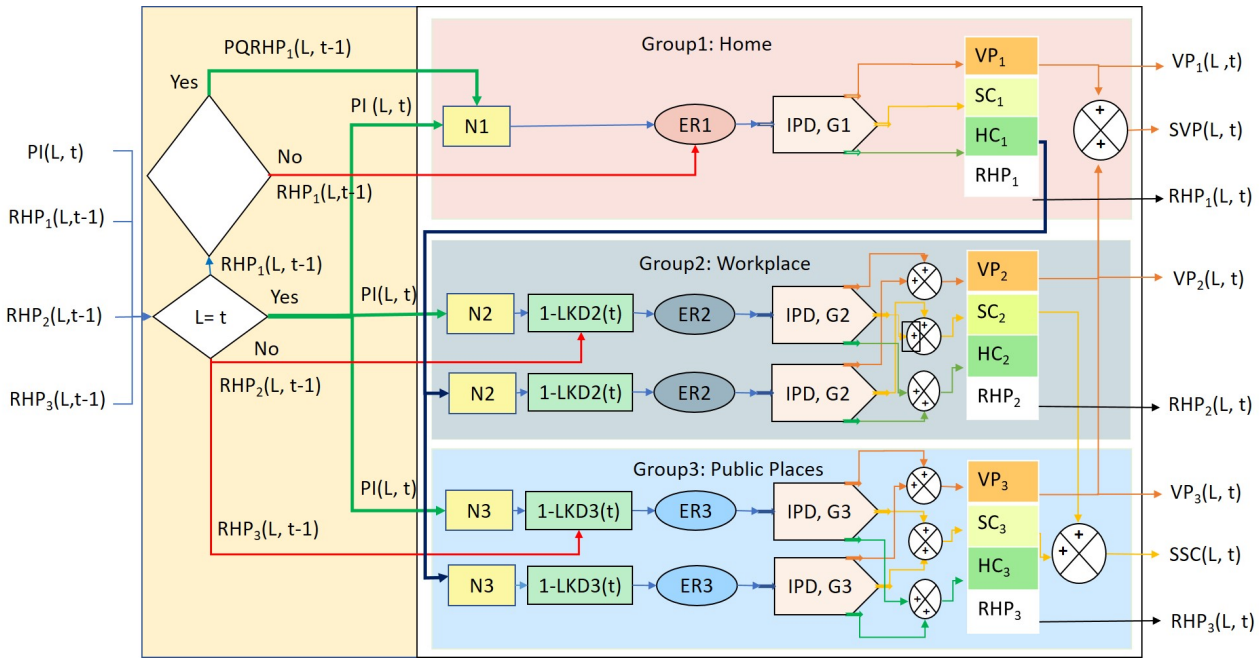



**Figure 3** Computational Block (CB)

### The Virus Proliferation Network

In the beginning, there are no COVID-19 cases. One person, E0, comes from outside. He happens to be a silent carrier of SARS-CoV-2. Thus, he is labeled as a Silent Carrier External, E0. As shown in Fig. 1, he meets his family, people at the workplace and public places. During this interaction, some of the people get infected and become carriers of SARS-CoV-2. During these group conversations, $SC_1$, $SC_2$, and $SC_3$ people are silent carriers who are infected but never show symptoms, and $VP_1$, $VP_2$, $VP_3$ are virus patients who start showing symptoms in few days. For example, in Fig. 1, when E0 meets people at different groups, P2, P7, P12, P16, P20 become silent carriers while P1, P5, P6, P10, P11, P19 become virus patients. These silent carriers and virus patients do not infect instantaneously, they are eventually distributed to get asymptomatic and symptomatic carriers. (See section 'Infectious Asymptomatic, Pre-symptomatic, and Symptomatic Patients'). When these asymptomatic and pre-symptomatic carriers go to their households, they start a new cycle of infection. This means that they start a new level inside the VPN. However, not all asymptomatic carriers created in each CB up to this time can infect others all the time. It is observed that they are infectious only during their incubation period and no virus carrier would continue to infect beyond 14 days. To account for this, we consider carriers created only in the last 14 levels of the previous day.

The silent carriers and virus patients that go to the vertical block, first go through the Infectious Patient Module where they are reduced and distributed. This is explained in detail in the section titled Infectious Asymptomatic, Pre-symptomatic, and Symptomatic Patients. One of the inputs to the Infectious Patient Module is SSC i.e., Sum of the Silent carrier of this CB which is calculated using equation (1) and further processed as shown in Fig 4.

$$SSC\,(L,t) = SC_1\,(L,t) + SC_2\,(L,t) + SC_3\,(L,t) \tag{1}$$

Computational blocks also have Remaining Healthy People, RHPs, as output. But they go only to a horizontal block. In the horizontal block, we have the same household members and work colleagues. MIMANSA keeps track of these individuals and finds out how many people have not yet been infected. A next horizontal block is equivalent to a new day in the same households, same workplace, and the same neighborhood. Whereas a vertical block means a new circle of one household, a workplace, and a neighborhood.

The output of a computational block is the sum of all silent carriers (SSC) coming out of the $IPD_1$, $IPD_2$, and $IPD_3$, VP coming out of $IPD_1$, $IPD_2$, and $IPD_3$, and remaining healthy people (RHP). The healthy carriers are not carried forward to the next day cycle since we have presumed that a fomite can infect only for a 24-hour duration.

Each CB is labeled as CB(Lt). For instance, CB12 means CB at level = 1, and time = 2. This output is given to the vertical and horizontal blocks of the next day as shown in Fig. 4. The output of CB11 is given to CB22 and CB12 where CB22 forms the vertical block while CB12 forms the horizontal group. This is how the virus spreads its network. In a horizontal block, it is

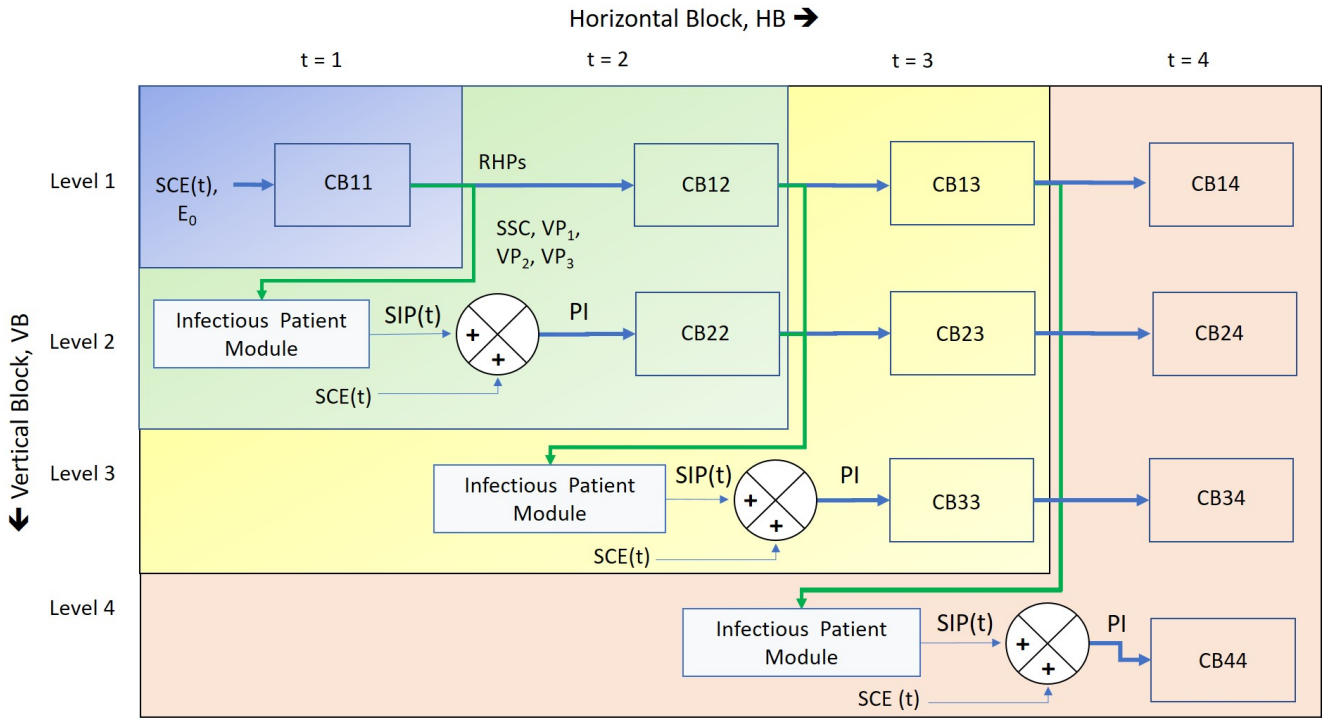



**Figure 4** Virus Proliferation Network (VPN)

the intra-group transmission while in a vertical block, it is the inter-group transmission. All computational blocks are integrated at the end to get an upper triangular matrix network.

### Infectious Asymptomatic, Pre-symptomatic, and Symptomatic Patients

In MIMANSA, the process of virus spread starts with the first silent carrier who comes from abroad, (Fig. 1) carrying the virus, meets some people at home, workplace, or public places, However, the people whom he meets do not necessarily become virus-infected patients that very moment. They either never show symptoms (asymptomatic), show symptoms (symptomatic), or show symptoms later (pre-symptomatic) during their incubation period but are infections during the course.

The Virus Patients, $VP_1$, $VP_2$, and $VP_3$ are distributed using the Weibull probability distribution, PD. The sum of all the previously distributed VPs going back up to 14 days, at time t, are symptomatic patients. These symptomatic patients come out as DNC as shown in Fig. 5 using equations (2), (3), and (4).

$$SYMVP_i(L,T) = VP_i(L,t) * PD(j) \tag{2}$$

where, T = t + j, i = 1, 2, 3, j varies from 0 to 13.

$$DNC(L,t) = SYMVP_1(L,t) + SYMVP_2(L,t) + SYMVP_3(L,t) \tag{3}$$

$$DNC(t) = \sum_{i=0}^{k2} DNC(L-i,t) \tag{4}$$

where, k2 = ((L-i) >0) and (i<=13)

The pre-symptomatic patients often start infecting 3 days before showing symptoms[22], therefore to calculate these pre-symptomatic patients (UVPs) the sum of all the distributed patients coming up in the next 3 days is taken. UVPs are calculated using equation (5) and processed further as shown in equations (6) and (7) as in Fig. 5. $UVP_1$ is also quarantined if one of the family members is detected as VP. More details are discussed in the section titled 'Quarantining'.

$$UVP_{i_1}(t) = \sum_{i=1}^{k1} \sum_{n=t+1}^{t+3} SYMVP_{i_1}(L,n) \tag{5}$$

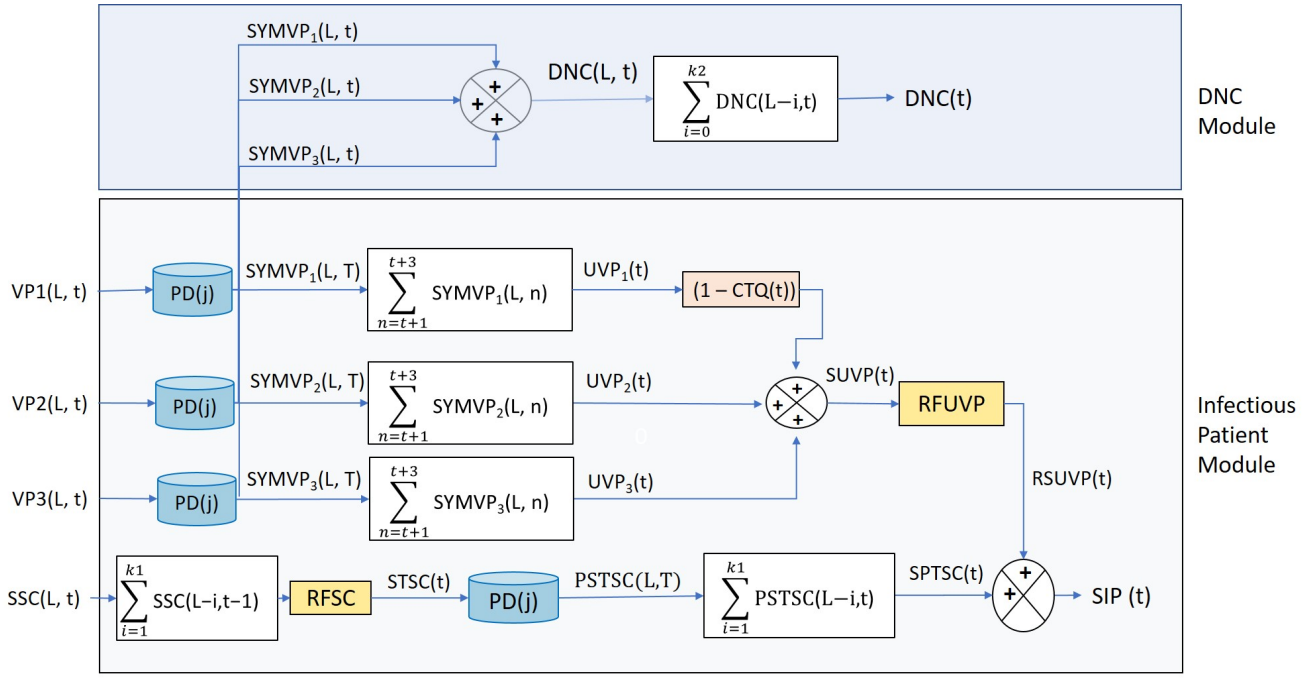



**Figure 5** Virus 'Daily New Cases' and 'Infectious Patient' module

where, $i_1$ = 1,2,3, k1 = ((L-i) > 0) and (i<=14)

$$SUVP(t) = UVP_1(t) * (1 - CTQ(t)) + UVP_2(t) + UVP_3(t) \tag{6}$$

$$RSUVP(t) = SUVP(t) * RFUVP \tag{7}$$

The asymptomatic patients are often infectious during their incubation period. So, the sum of the last 14 levels of SSC is taken into consideration and distributed using the PD. These distributed SSCs, at the time t, are summed up to get SPTSC as per equations (8), (9), and (10) and as shown in Fig. 5.

$$STSC(t) = RFSC * (\sum_{i=1}^{k1} SSC(L-i, t-1)) \tag{8}$$

$$PSTSC(L,T) = STSC(t) * PD(j) \tag{9}$$

$$SPTSC(t) = \sum_{i=1}^{k1} PSTSC(L-i,t) \tag{10}$$

where, T = t + j, j varies from 0 to 13, k1 = ((L-i) > 0) and (i<=14).

However, these pre-symptomatic and asymptomatic patients do not infect everybody they meet. It is observed that only a certain percentage of people who meet pre-symptomatic or asymptomatic patient gets infected which in turn can transmit the virus further. Thus, SSCs are reduced by a factor called RFSC (Reduction Factor for Silent Carriers) and the secondary attack rate of the pre-symptomatic patients is called the Reduction Factor for Unidentified Patients, RFUVP.

$$SIP(t) = SPTSC(t) + RSUVP(t) \tag{11}$$

$$PI(t) = SIP(t) + SCE(t) \tag{12}$$

The sum of these reduced UVPs and SPTSC are the total infectious people i.e., Sum of Infected Patients (SIP). These are calculated using equations (11). These SIPs as shown in equation (12) are added with SCEs and given as input to the computational block.

## Quarantining

Contact Tracing and Quarantining have been used in many countries to control the spread of COVID-19. In this method, if a patient tests positive then all people s/he has met in the last 14 days are quarantined. Contact tracing, in general, is an expensive option as it is difficult to trace all the people with whom the infected person has interacted. Therefore, we suggest household quarantining as an alternative to Contact tracing and quarantining or (CTQ). In this method, if a patient is tested positive, his family is quarantined at home. This reduces further virus spread since the people in contact with the patient are not allowed to mix with other people.

CTQ factor used to simulate the household quarantining, ranges from 0 to 1. It is applied only to group 1 of the model, that is, the household group. CTQ affects the silent carriers (SC), Remaining Healthy People (RHP), and Unidentified Virus Patients (UVP) from group 1.

When quarantining is applied to silent carriers, we first calculate the Individual Probability of Silent Carriers from group 1 ($IPSC_1$). $IPSC_1$ is nothing but the number of silent carriers that start infecting a particular computational block. $IPSC_1$ is calculated by using the silent carrier distribution of each level and up to time equals level. The silent carriers to be quarantined (QPSC) are equal to $IPSC_1$.

Ideally, the silent carriers infecting at any computational block are proportional to $IPDFVP_1$, that is, virus patients infecting at that computational block.

Therefore, if SCR is greater than IR, and $IPSC_1$ are very high then we presume silent carriers are from different families. In this case, QPSC is proportional to the virus infecting patients in that computational block and it is calculated using the equation (13).

$$QPSC(L,T) = (SCR/IR) * SYMVP_1(L,t) \tag{13}$$

Once we find out the number of silent carriers to be quarantined, we multiply them by the CTQ parameter and subtract them from silent carriers from that computational block. As the CTQ parameter tells us the percentage of quarantining that is being followed, multiplying QPSC with CTQ will give us the exact number of silent carriers that are quarantined. After subtracting these from silent carriers of group 1, we get the final number of silent carriers that will infect even though quarantining policy is applied.

Similarly, we also reduce $RHP_1$, we have to quarantine this $RHP_1$. If these $RHP_1$ are too high, we find $RHP_1$ to be quarantined by finding the $RHP_1$ proportional to $SYMVP_1$ by using the RHP rate which is calculated using equation (14).

$$RHP\ Rate = 1 - IR - SCR \tag{14}$$

$$QRHP(L,T) = ((RHP\ Rate)/IR) * SYMVP_1(L,t) \tag{15}$$

$$PQRHP_1(L,t) = RHP_1(L,t) - CTQ(t) * QRHP(L,t) \tag{16}$$

Just like silent carriers, we multiply these QRHP by CTQ to get the exact number of $RHP_1$ to be quarantined and then subtract it from $RHP_1$ to get the final number of $RHP_1$ that infects others. These are calculated using equations (15) and (16).

In the case of Unidentified Virus Patients (UVP), we also reduce UVP from group 1. But here, we simply multiply $UVP_1$ by CTQ to get the number of quarantined UVP (QUVP). The UVPs, RHPs, and $SC_1$s are reduced only if a virus patient is detected in the household i.e., VPPD > 0, and Quarantining policy is applied.

### Mathematical Outcome of the Computational Block

Mathematical equations for VP, SSC, RHPs when all the three groups are put together, are shown below. The following equations are per day rates or difference equations where the duration of the difference is one day. Thus, all values are daily values. Let,

$$\alpha = N2 * (1\text{-}LKD2\ (t)) * ER2\ (t),$$
$$\beta = N3 * (1\text{-}LKD3\ (t)) * ER3\ (t),$$
$$\gamma = PI\ (t) * N1 * ER1\ (t) * SCR\ (t),$$
$$\delta = PI\ (t) * N1 * ER1\ (t) * IR\ (t),$$

Then,

$$VP(L,t) = \delta * [1 + (\alpha + \beta) * HCR] + (PI(t) * IR(t) * (\alpha + \beta)) \tag{17}$$

$$SC(L,t) = \gamma * [1 + (\alpha + \beta) * HCR] + (PI(t) * SCR(t) * (\alpha + \beta)) \tag{18}$$

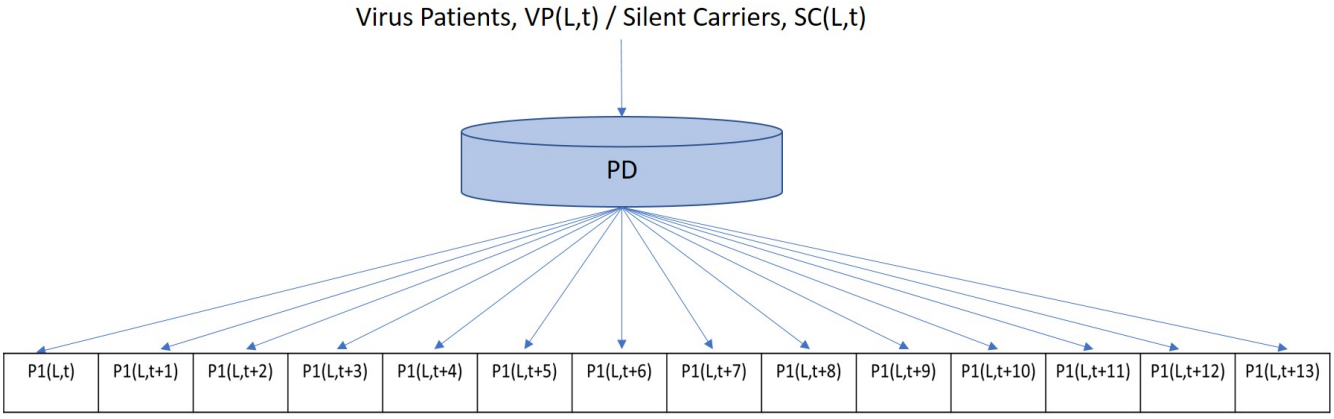
$$RHP_1(L,t) = (PI(t) * N1 * ER1(t)) * [1 - (IR(t) + SCR)(t)] \tag{19}$$

$$RHP_2(L,t) = \alpha * [PI(t) + ((\delta / IR(t)) * HCR)] * [1 - (IR(t) + SCR(t))] \tag{20}$$

$$RHP_3(L,t) = \beta * [PI(t) + ((\delta / IR(t)) * HCR)] * [1 - (IR(t) + SCR)(t)] \tag{21}$$

These equations get a bit more complex once probability distribution is applied on VP (L, t) and SSC (L, t).

### Simulating the Incubation Period by Probability Distribution




| Level, L | time, t | VP/SC | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | | |
| | | WD(T) | 2.39E-06 | 0.000229 | 0.003751 | 0.025529 | 0.10708 | 0.288617 | 0.404663 | 0.164304 | 0.00582 | 2.04E-06 | 1.60E-13 | 1.11E-26 | 3.38E-50 | 5.66E-89 | | |
| 1 | 1 | 10 | 0.0000 | 0.0023 | 0.0375 | 0.2553 | 1.0708 | 2.8862 | 4.0466 | 1.6430 | 0.0582 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | | |
| 1 | 2 | 20 | | 0.0000 | 0.0046 | 0.0750 | 0.5106 | 2.1416 | 5.7723 | 8.0933 | 3.2861 | 0.1165 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | |
| 2 | 2 | 30 | | 0.0001 | 0.0069 | 0.1125 | 0.7659 | 3.2123 | 8.6585 | 12.1399 | 4.9291 | 0.1747 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | |
| 1 | 3 | 40 | | | 0.0001 | 0.0092 | 0.1500 | 1.0211 | 4.2831 | 11.5447 | 16.1865 | 6.5722 | 0.2330 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | 3 | 50 | | | 0.0001 | 0.0114 | 0.1876 | 1.2764 | 5.3539 | 14.4309 | 20.2332 | 8.2152 | 0.2912 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3 | 3 | 60 | | | 0.0001 | 0.0137 | 0.2251 | 1.5317 | 6.4247 | 17.3170 | 24.2798 | 9.8583 | 0.3494 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Figure 6** Distribution of virus patients or silent carriers during their incubation period

The SARS-CoV-2 virus goes through a 14-day incubation cycle. A person infected with the virus may show symptoms anytime within 14 days from the time of infection[23]. Given a group of 100 people infected at the same time how many would be symptomatic on a given day is simulated using a Weibull probability distribution, WD(T) with shape 7.7 and scale 7 (PD) in MIMANSA[24]. This Weibull distribution is mentioned in the third row of Fig. 6 while the distribution of the count of patients after the infection is depicted in the next consecutive rows. . Additionally, the second row is labeled as time, and the columns are labeled as Level, time, and VPs/SCs. As shown in Fig 6, let us consider an example where 10 virus patients are generated in CB11, 20 in CB12, 30 in CB22, and so on. These values are distributed using the Weibull distribution. It is observed in row four that on day 1 for CB11 only one person shows up with symptoms on t = 5, two people on t = 6, and so on. As time changes, there is a shift in the starting point of each distribution. For instance, when a person is infected on day 2, their distribution is shifted by one day as their incubation period starts from day 2 and its 14-day period cycle will be over on day 15. This indicates a staggering of individuals who show up with symptoms. Now when we have to count the number of cases on a particular day, we have to column-wise add all patients that showed symptoms on day d and go back 14 levels from the current level.

## Transmissibility

The SARS-CoV-2 constantly mutates itself to become strong and survive by either becoming more transmissible, invading the immunity of the individuals, or both. Variants like delta, alpha, gamma, and beta are increasingly prevalent across the globe. In our study, we have only considered the variants of concerns (VOCs) with prevalence greater than 10% in a region. To account for the transmissibility of a heterogeneous group of SARS-CoV-2 variants, a weighted average of prevalence ($\rho$) of multiple variants and its transmissibility ($\tau$) is calculated. The transmissibility in the heterogeneous virus population is calculated as per equations (22) and (23).

$$\rho_{nv}(t) = 1 - \sum_{i=1}^{n} \rho_i(t) \tag{22}$$

$$\tau(t) = \sum_{i=1}^{n} \rho_i(t) * \tau_i(t) + \rho_{nv}(t) * \tau_{nv}(t) \tag{23}$$

where,
$\rho_{nv}$ = Prevalence of the baseline virus
$\rho_i$ = Prevalence of a specific variant
$\tau$= Weighted average transmissibility
$\tau_i$ = Transmissibility of a specific variant
$\tau_{nv}$ = Transmissibility of the baseline virus

The VOCs used for this study for the United States of America (USA), United Kingdom (UK), and California (CA) and their transmissibility range[25] is given in Table 2. The presence of a variant in a given region is denoted by Yes while No represents its absence in Table 2.

For simplicity and ease in model training, we have used the minimum value of the transmissibility range for our analysis. It is assumed that the variants and sub-lineages have similar transmissibility.

| Variant | Baseline Variant | B.1.617.2 | B.1.1.7 | AY.4 | AY.25 | P.1 | AY.3 |
|---|---|---|---|---|---|---|---|
| USA | Yes | Yes | Yes | No | No | Yes | Yes |
| CA | Yes | Yes | Yes | Yes | Yes | Yes | No |
| UK | Yes | Yes | Yes | Yes | No | No | No |
| $\tau$ | 1 | (2.61, 3.77) | (1.41, 1.63) | (2.61, 3.77) | (2.61, 3.77) | (1.39, 2.19) | (2.61, 3.77) |

**Table 2** Variants and its transmissibility

## Vaccination

With an increasing number of cases and new variants emerging each day, there is a need for vaccines to control the spread of COVID-19. Having rolled out vaccines, we have people who are partially vaccinated with only a single dose ($\sigma_1$), fully vaccinated ($\sigma_2$), and others who remain unvaccinated ($\sigma_0$). Since vaccine efficacy is not 100%, some vaccinated people remain unprotected despite taking the vaccine. Let the efficacy of the vaccines for the single-dose category be $\varepsilon_1$, and for double dose be $\varepsilon_2$. Thus, after vaccination, some vaccinated people are protected. Those with a single-dose and protected are labeled as $\sigma_{11}$,

while fully vaccinated and protected are called $\sigma_{21}$. Others, despite taking the vaccine, remain unprotected ($\sigma_{12}$, $\sigma_{22}$) depending upon the efficacy of the vaccines for partial ($\varepsilon_1$) and complete ($\varepsilon_2$) doses. The unvaccinated people, partially protected, and completely protected are modeled using equations (24), (25), (26), (27), and (28).

$$\sigma_0(t) = 1 - (\sigma_1(t) + \sigma_2(t)) \tag{24}$$

$$\sigma_{11}(t) = \sigma_1(t) * \varepsilon_1 \tag{25}$$

$$\sigma_{12}(t) = \sigma_1(t) * (1 - \varepsilon_1) \tag{26}$$

$$\sigma_{21}(t) = \sigma_2(t) * \varepsilon_2 \tag{27}$$

$$\sigma_{22}(t) = \sigma_2(t) * (1 - \varepsilon_2) \tag{28}$$

Where, $\sigma_{xy}$ denotes the number of partially or fully vaccinated protected people based on the efficacy of the vaccine.

x denotes fully or partially vaccinated people. Here, if x is 1 it represents partially whereas x equal to 2 represents fully vaccinated people.

y denotes the protected vaccinated people. The value 1 denotes protected vaccinated people while the value 2 stands for unprotected vaccinated people.

### Reduction of Mask Usage Ratio

Mask Usage is commonly referred to as Mask Usage Ratio or MUR in our study. Some countries have allowed to not wear a mask for vaccinated people. However, because vaccines are not 100% effective and some unvaccinated people also have stopped using masks[26,27], we need to reduce the observed mask usage. The reduction in mask usage is only for unprotected vaccinated individuals and unvaccinated individuals i.e., $\sigma_{12}$, $\sigma_{22}$, and $\sigma_0$ respectively as only these people are responsible for most of the further transmission of the virus while transmission due to vaccinated individuals is comparatively low. This effect is modeled using equations (30) and (31). The reduction in MUR is denoted as Mask Usage Ratio Reduction or MURR.

Let $\mu$represent the number of people not wearing a mask then it is modeled as a Gompertz equation as shown in equation (29). The suffix 0, 1, and 2 showcases unvaccinated, partially, and fully vaccinated, people who no longer wear a mask.

$$\mu(t) = c_1 * e^{-c_2 * e^{-c_3 * t}} \tag{29}$$

$$MURR(t) = \sigma_{12}(t) * \mu_1(t) + \sigma_{22}(t) * \mu_2(t) + \sigma_0(t) * \mu_0(t) \tag{30}$$

$$MUR(t) = MUR(t) * (1 - MURR(t)) \tag{31}$$

The parameter values $c_1$, $c_2$, and $c_3$ for $\mu_0$, $\mu_1$, and $\mu_2$ are modeled based on some surveys. More details are discussed in Data, Parameters, and Parameter Identification section.

### Reduction of IR and SCR

The probability of transmitting the virus to other individuals reduces because of vaccination, especially due to protected vaccinated individuals. Thus, we reduce the infection rate, IR, and the silent carrier rate, SCR, in proportion to the number of partially or fully protected vaccinated people. However, the reduction is not linear, as more people are vaccinated, more antibodies are created among the vaccinated individuals, and the probability of transmitting the virus reduces even more. This effect is modeled using the Vaccination Impact factor ($\varkappa$) and equation (32), and (33).

$$IR(t) = IR(t) * (1 - \kappa * (\sigma_{11}(t) + \sigma_{21}(t))) \tag{32}$$

$$SCR(t) = SCR(t) * (1 - \kappa * (\sigma_{11}(t) + \sigma_{21}(t))) \tag{33}$$

## Data, Parameters, and Parameter Identification

In this study, we have developed a model to study the number of cases of COVID-19 for the USA, California, and the UK. The total system diagram involves multiple parameters as well as data. The following is the description that outlines details on data, parameters, and model training for parameter identification.

1. Data for daily COVID-19 cases
   The daily USA COVID-19 cases from the OWID website[28] are used to train the MIMANSA model from February 15, 2020, to July 31, 2021, which is later validated for the next 15 days i.e., August 15, 2021. The UK and CA are trained for the same time frame. The actual data for the CA is taken from LA times[29] while the UK's actual daily number of COVID-19 cases is taken from the UK.GOV website[30].

2. Lockdown data and its usage
   Community Mobility reports (CMR) website[31] give an insight into changes in the traveling habits of the people due to changes in government policies like lockdown. It shows the trend in the movement of people in different places like parks, transit stations, workplaces, and residential areas at different times. In our study, we have considered the effect of lockdown at residential places (LKD1), workplaces (LKD2), and transit stations (LKD3). If the lockdown parameter (LKD2) is 0.6, it means that the lockdown was 60% adhered to. For instance, in case the percentage change of people's mobility at transit stations on a day is -73%. It shows a 73% decline in people using transit stations i.e., LKD3 = 0.73. Any increase in similar cases is considered zero. However, in the case of residential places, the interpretation of lockdown is different. Unlike LKD2 and LKD3, LKD1 represents the change in the duration of people's stay at home. These changes will be only 50% as 50% of the time people were already at home so that doesn't change much. For example, when we say the change in the duration of people's stay is 14%, it is the change in 50% of the time where people would ideally go to the office but stayed back at home due to lockdown. Therefore, to incorporate the effect of the whole day, LKD1 is calculated as 2*(14/100) = 0.28. Any decreasing changes are considered as zero as these do not reflect the effect of lockdown policies.

3. Mask Usage data
   The mask usage graphs as a function of days are given on the IHME website[32]. The graphs were digitized to get the mask usage series for every dataset. This data is also known as Mask Usage ratio (MUR) is used in the simulations.

   It is observed that masks do not give complete protection against the virus. The best type of mask is the N-91 that prevents the SARS-CoV-2 transmission by 95% when used properly[21]. This percentage starts going down due to the type of mask, and its proper usage. Additionally, the mask usage value depends on the enclosed space, the number of people meeting, enclosed or closed space, duration of contact, social distancing, and last but not least is the reliability of the answers given during the survey. Thus, during the parameter estimation stage, we reduce the mask usage by 30

   In the case of the USA, no mask day was announced on May 14, 2021, and declared that fully vaccinated individuals need not wear a mask and can return to normalcy. However, there are a certain set of partially and unvaccinated people who also plan to not wear a mask. As per the survey conducted by Kaiser Family Foundation, it is observed that 19% are fully vaccinated, 9% partially and 26% of unvaccinated people no longer plan to wear a mask[26]. Another survey showcases the number of people no longer wearing a mask in the weeks of April 19 and May 3, 2021[27]. Using these two data we modeled the number of people who no longer wear a mask using the Gompertz equation as shown in equation (29).

   The parameters, $\mu_i$ = [c1, c2, c3]

   where i = 0 to 2, c1, c2, c3 are parameters in the equation (29)

   $\mu_0$ = [0.26, 1.4, 0.05],
   $\mu_1$ = [0.09, 2.5, 0.075],
   $\mu_2$ = [0.19, 3, 0.037]

   As it can be seen, a very small percentage of people have already stopped using masks even before no mask day was announced[27] and thus, we start reducing the mask usage with the onset of vaccination. We assume a similar situation for the number of people not wearing a mask in California.

   In the case of the UK, Freedom Day was declared on July 19, 2021, in England wherein the legal requirement to wear a mask ended. However, masks must still be worn in Wales and Scotland. In Northern Ireland, masks still must be worn at public transport, shops, and hospitals while it is not compulsory in places of worship and for students in school classrooms[33]. As per a study, despite the lifting restriction on mask usage, almost 64% of people were planning to wear a

mask in transports and 60% are planning to avoid crowded places in England[34]. All in all, we can say that majority of the people continue to wear a mask, and following this, we have assumed in our study that the majority of the people in the UK continue to wear a mask despite the ease in some restrictions.

4. Number of people in a group
   The number of people in each of the 3 groups is selected as follows. In general, it is reasonable to assume that a household consists of 4 people. At the workplace, one meets a limited number of people compared to in public places. As per Del Valle[20], one meets 22 people every day. With that total in mind, we select 4 people at home, 10 people at the workplace, and 8 people in public places. Although these numbers may vary in individual cases, on average, it appears to be a reasonable distribution.

5. Regional Environmental Variability Factor
   MIMANSA is a model for simulating the spread of SARS-CoV-2 globally. The underlying equations remain the same all over the world. However, it is known that the spread of any virus depends on the environment in each region. Besides climatic conditions, people have different social norms and behavior changes as per the norms. All these factors impact a given region. To account for this variability, we have considered a parameter called Regional Environmental Variability (REV). The REV impacts the exposure rate, silent carrier rate, infection rate, and household quarantining.

6. Daily Environmental Variability Factor
   Every day, the exposure to SARS-CoV-2 in a region may change due to holidays, festivals, games, regional family events, etc. These variations are captured by this parameter. This is the only parameter that one has to adjust after the model is trained.

7. Exposure and Infection rates
   All exposure rates, the silent carrier rate, and the infection rate were modeled as a function of the lockdown and mask usage ratio. ER1 is dependent only on LKD1 because people do not wear a mask at home while ER2 and ER3 are dependent on both, the lockdown values, and the non-mask user ratio (NMUR).

   It is observed that 30.8% of infections are caused because of meals gathering, 30.13% because of household contact while community transmission, and workplace results in 18.59% and 19.87% infections[35]. The ratio of household/workplace i.e., 1.5163, and community transmission/workplace which is 0.9355, gives an insight into the proportion of infection that is likely to occur at home and travel places depending upon the workplace environment in various regional environmental conditions. Thus, the regional environmental variability factor for ER1 (REV_ER1), and ER3 (REV_ER3) are calculated as per equations (34) and (35).

$$REV\_ER1 = 1.5163 * REV\_ER2 \tag{34}$$

$$REV\_ER3 = 0.9355 * REV\_ER2 \tag{35}$$

   The infection rate and silent carrier rate is dependent only on mask usage since it depends on interpersonal interaction. The Healthy Carrier Rate (HCR) is taken low since the possibility of fomite transmitted infection is low. We assumed a constant value of HCR to be 0.1.

8. Virus Transmissibility
   The exposure rate, silent carrier rate, and infection rate get affected by the transmissibility of the virus. Higher the transmissibility, the higher would be the exposure, silent carriers, and symptomatic patients.

   Data used for calculating the transmissibility is obtained from the outbreak website[36]. The prevalence data of various variants are first digitized and later weighted average transmissibility for both the countries and a state is calculated using equations (22), and (23).

9. Vaccinated population
   Vaccinated people are better protected against the SARS-CoV-2 virus. This reduces the number of possible hosts for the virus. The impact of this is considered in the model by the equations (30), (31), (32), and (33). The vaccine's efficacy for USA's and CA's vaccines for partially and fully vaccinated people is observed as 73% and 93% while that of UK's vaccines is 64.1% and 70.3%[37]. The time-series data of the percentage of people getting vaccinated in the USA and the UK are taken from OWID[28] while California data is taken from USAfacts[38,39].

10. Household Quarantining
    Here, to get the CTQ values, it is assumed that the minimum number of people who would abide by the quarantining policy is proportional to the number of people wearing a mask. However, there is also a possibility that more people follow this practice due to the government's strict enforcement. Taking these things into account, CTQ value varies with MUR as an independent constant. This constant depends on social behavior in a region. In some areas, people may abide by it more than using a mask whereas in some areas it could be the same number or less compared to the mask usage. Thus, this regional environmental variability in the case of household quarantining (REV_CTQ) is adjusted during the training phase and household quarantining is modeled as shown in equation (36).

$$CTQ(t) = REV\_CTQ * MUR(t) \tag{36}$$

11. Secondary Attack Rate
    It is observed that not all pre-symptomatic and asymptomatic patients are responsible for the transmission of SARS-CoV-2. The secondary attack rate for pre-symptomatic (RFUVP) and asymptomatic (RFUSC) are 26.7% and 18.8% respectively[40,41]. Therefore, we reduce them by these reduction factors before estimating the infectious population for the next level. Moreover, it is also observed that only 2% of people are responsible for 90% spread in the society[42].

12. Training Model
    MIMANSA is trained using the data of two countries the USA and the UK, and one state, California. This will help understand the variations in the system parameters among different countries and a country and state. The actual number of COVID-19 cases data for the USA, California, and the UK is taken from the OWID website[28], LA times[29], and UK.gov[30] websites The lockdown data was derived from the Community Mobility reports[31]. The mask usage is found from the map published by the Institute for Health Metrics and Evaluation (IHME), an independent global health research center at the University of Washington[32]. The model is trained from February 15, 2020, to July 31, 2021, using MIMANSA. MIMANSA mimics the real-world observations to simulate the COVID-19 spread using parameters like lockdown, mask usage, household quarantining, exposure rate, etc. In the case of the USA, CA, and the UK, the parameters are adjusted per the data available on websites[28–32]. Initially, the parameters like exposure rates, silent carrier ratio, and infection rate, were manually adjusted as per our understanding and observation. While training the MIMANSA, a strong correlation was observed among these adjusted values of ERs, SCR, IR, MUR, and LKD. Later, the exponential equations were modeled using data modeling techniques for these adjusted values. These equations were later fine-tuned to match the actual data. Equation (38) is used to train for the countries and a state.

    We define the No Mask Usage Ratio (NMUR) as the percentage of people who are not using masks. This is calculated using equation (37).

$$NMUR(t) = 1 - MUR(t) \tag{37}$$

$$X(t) = \rho(t) * DEV(t) * REV(t) * (a_1 * e^{-LKD_i(t) * l_1} + a_2 * e^{-NMUR(t) * (l_2)} - c) \tag{38}$$

    $X_i$ is the exposure rate for group 1, 2 and 3 (ER1, ER2, ER3), Silent Carrier Rate (SCR), or Infection Rate (IR). Parameter set, $S[X_i] = [REV, a_1, l_1, a_2, l_2, c]$.

    The detailed parameter values for each parameter set are mentioned in Table 3. Parameters tuned during the training phase include the Regional Environmental Variability impacting the Exposure Rate, Silent Carrier Rate, infection rate, and Quarantine ratio. The other parameter includes the Vaccination Impact factor, $\varkappa$, Additionally, the Daily Environmental Variability parameter was also tuned for both country and the state during the training period.

    The trained model is also evaluated using a standard statistical measure, the Root Mean Square Error, RMSE.

### Validation and Prediction

The trained MIMANSA model for USA data is validated from August 1, 2021, to August 15, 2021. The same time period is used for the UK as well as California. Predictions are made as per the availability of the actual number of cases data. Projections for the USA and UK are made till September 6, 2021, while CA has projections till September 2, 2021. The actual data of MUR, LKD, transmissibility, vaccination are used to validate and predict the number of cases.

| | REV_ER2 | REV_SCR | REV_IR | REV_CTQ | $\varkappa$ |
|---|---|---|---|---|---|
| USA | 0.64 | 0.7 | 0.48 | 1.33 | 1.2 |
| CA | 0.63 | 0.7 | 0.48 | 1.35 | 1.26 |
| UK | 0.5 | 0.62 | 0.48 | 1.36 | 1.042 |

**Table 3** Training Parameter Values

### Effect of Vaccination

Vaccines have been an integral part of controlling the COVID-19 spread. To see the effect of the absence of vaccination, the number of people who received vaccines was set to zero. All other parameters were kept the same as they were earlier. However, in reality, if there is an increase in the number of COVID-19 cases, policies are often made to reduce mobility, and people too are more likely to wear masks. To account for this, the increasing lockdown trend, seen historically, was replicated to reduce the sudden increase in the cases due to the absence of vaccination. The mask usage was kept as it was since the mask usage values could not be increased beyond a certain limit.

## Results and Analysis

The trained model, validation, and predictions for the USA, CA, and the UK are shown in Fig. 7, Fig. 8, and Fig. 9 respectively. It is observed that the trained model compares well with the actual data. The RMSE values for these fitted data for the USA, CA, and the UK are 247340.78, 32165.91, and 43169.12. The average percentage error in the validation stage is around 0.30%, 1.84%, and 1.06% for the USA, CA, and the UK respectively.

The Daily Environmental Variability, DEV, values used to tune the MIMANSA model while training the USA, CA, and the UK are shown in Fig. 10. The effect of reducing the vaccination data to zero and later increasing the lockdown and the mask usage is shown in Fig. 11. The actual and increased lockdown data used to see this effect of vaccination on the number of cases is shown in Fig. 12.

## Discussion

#### *Simple vs complex model*

By now the literature is replete with multiple models for simulating COVID-19. Some are simple to use while others are more complex. At times, people feel that all models are made equal. If so, then why go for a complex model like MIMANSA? However, there is an obvious catch. Their thinking is that if a simple model can 'do the job', why bother to have a complex model. The key phrase in this is 'do the job'. One needs to define what that job is. Most of the time, the job is not just coming up with a number as a prediction. The job is to improve understanding, simulate complex scenarios, come up with answers to multiple different 'what if' scenarios. None of this is possible with a logistic, Gompertz, ARMA, and many more models that are simple to use. The majority of the models in the literature do not consider one or more of the parameters like regional variation, daily event variation, prevalent variants, transmissibility, mask usage, mobility, etc.

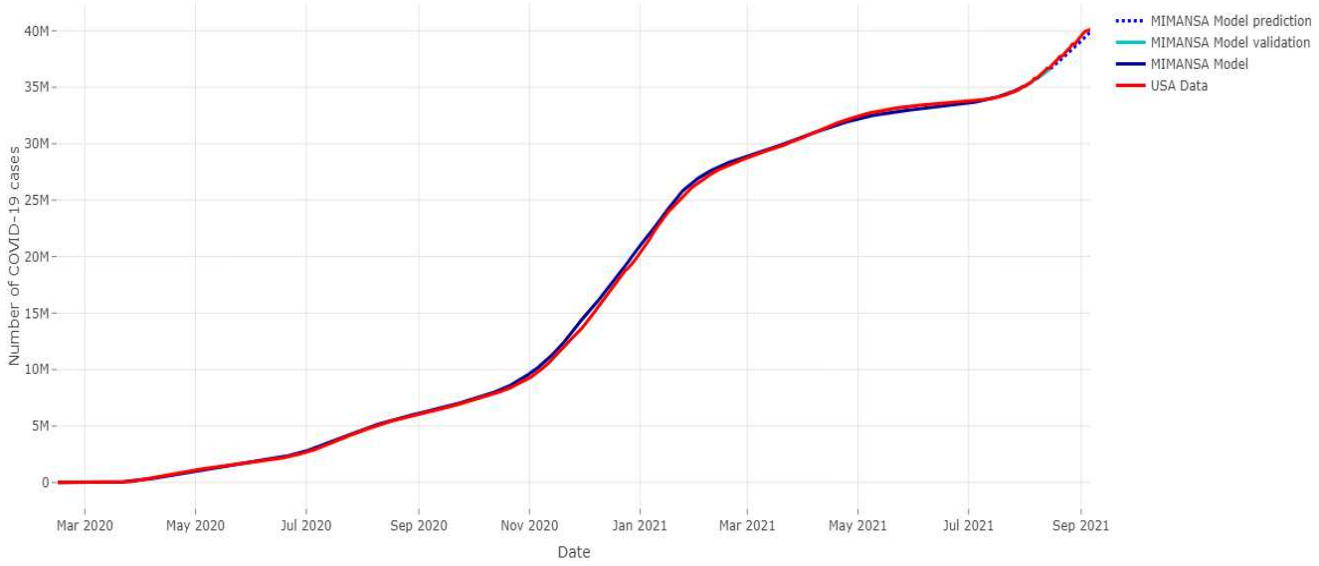
#### *Parametric Reduction*

As with any parametric model, our model has many parameters. The main difference between a standard parametric model and MIMANSA is that each of the parameters in MIMANSA is linked to observations with a significance in epidemiology. It is possible to club some parameters, however, it may lose the advantage of simulating each sub-system for better understanding.

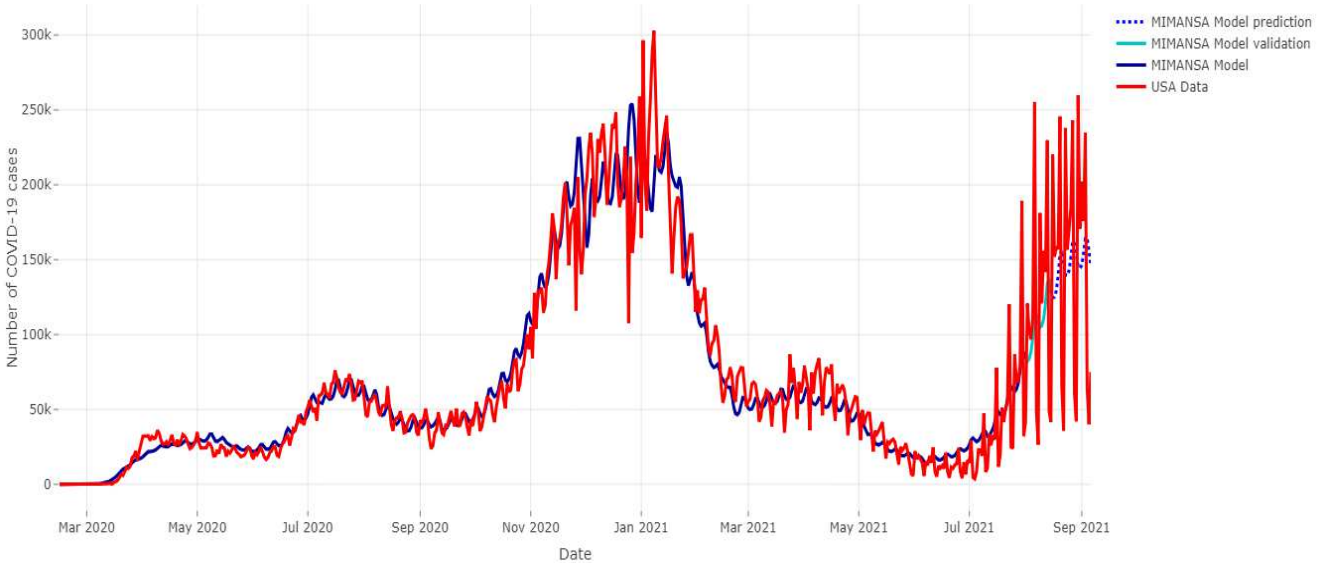
We achieved parametric reduction by going through the literature and replacing many parameters with constants backed up by data. Earlier mask usage was an educated guess and it was a parameter. Now that it is available on the internet, we use the time series for mask usage. It is the same case with the lockdown. There is a long list of parameters and variables that became constants or a time series based on the data available in the literature. For example, lockdown values were initially adjusted for a specific duration of time. Later, this mobility data was available in the literature and thus it was removed as an adjustable parameter. Similarly, now we have the data for virus transmissibility, mask usage, etc.

#### *Parameters required for simulation*

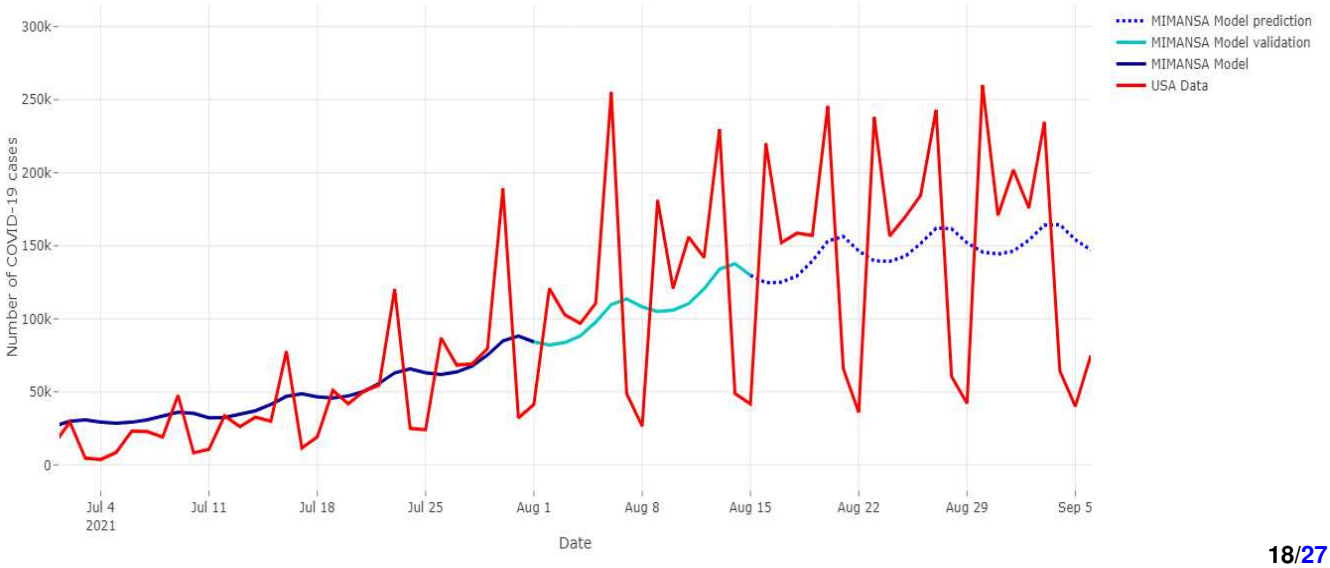
During the training phase, the model has five parameters. Parameter estimation for a given region is carried out during training. Once the training is completed, one may need to adjust for daily events such as a surge due to a football game, pubs being open for extended periods, family get-togethers due to festivals, etc. Such adjustments are possible by only adjusting one parameter called 'Daily Environmental Variability'.




**(a)** MIMANSA Training, Validation, Prediction for US cases (Cumulative)



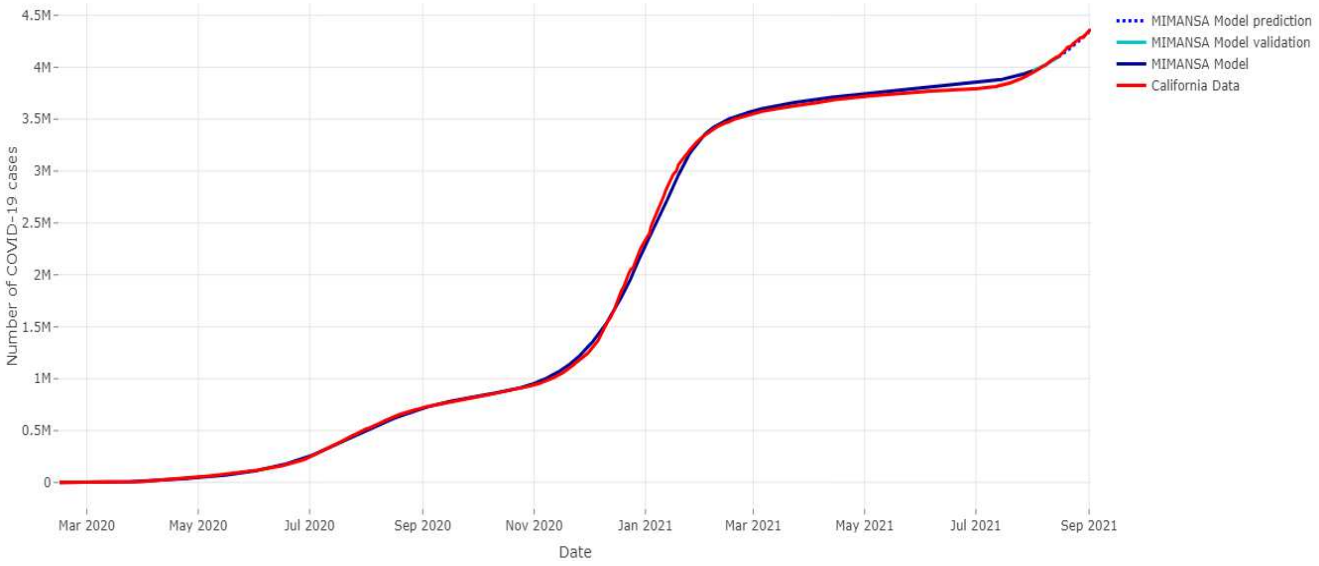

**(b)** MIMANSA Training, Validation, Prediction for US cases (Daily)



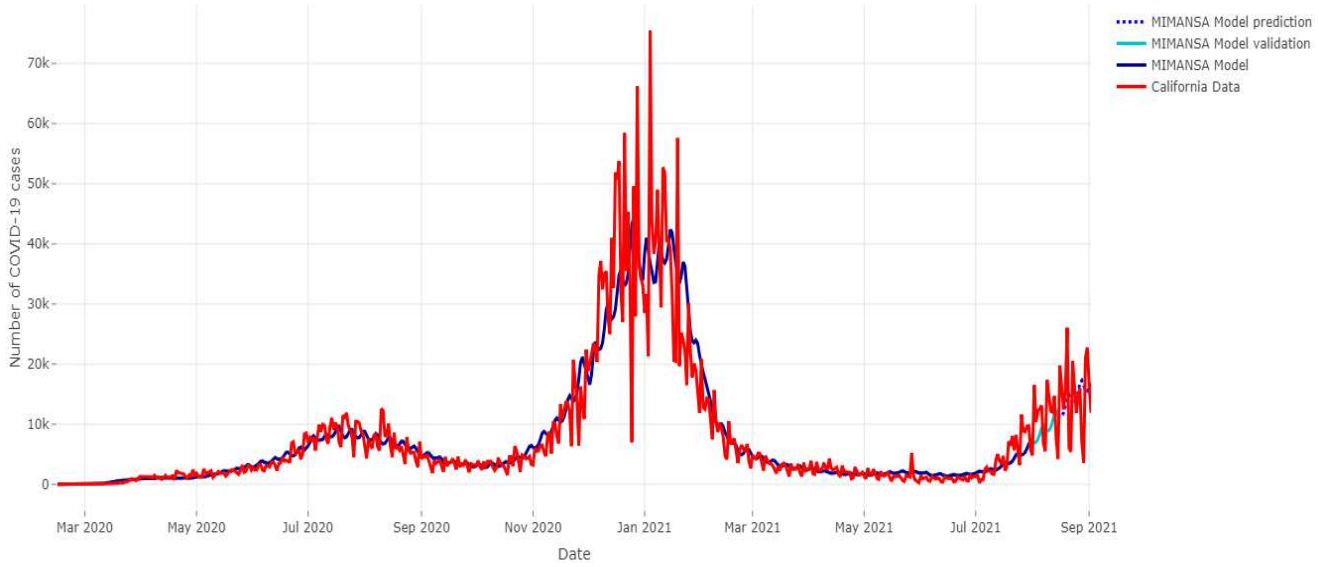



**(c)** MIMANSA Training, Validation, Prediction for US cases (daily from July 1, 2021)

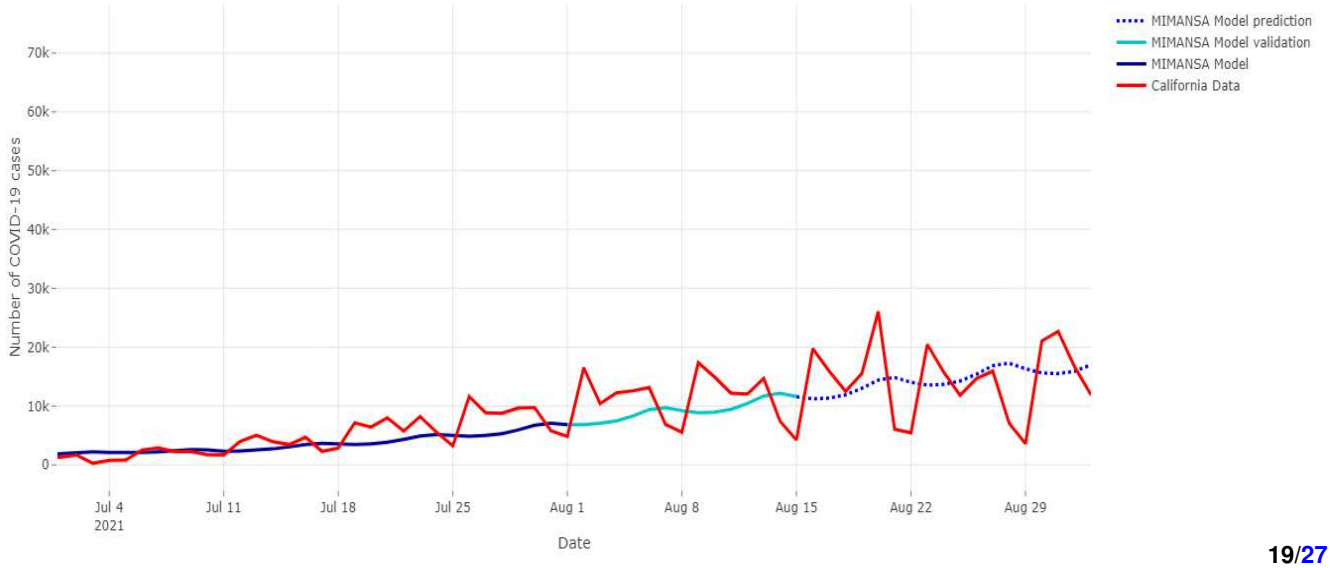
**Figure 7** United States




**(a)** MIMANSA Training, Validation, Prediction for California cases (Cumulative)



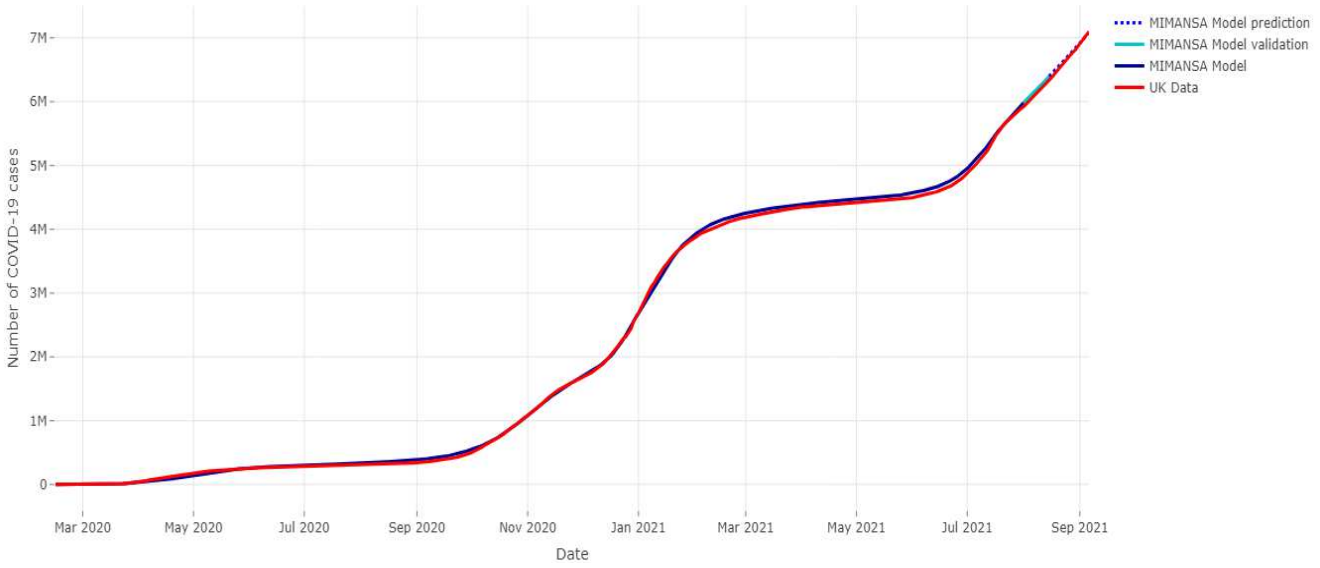

**(b)** MIMANSA Training, Validation, Prediction for California cases (Daily)



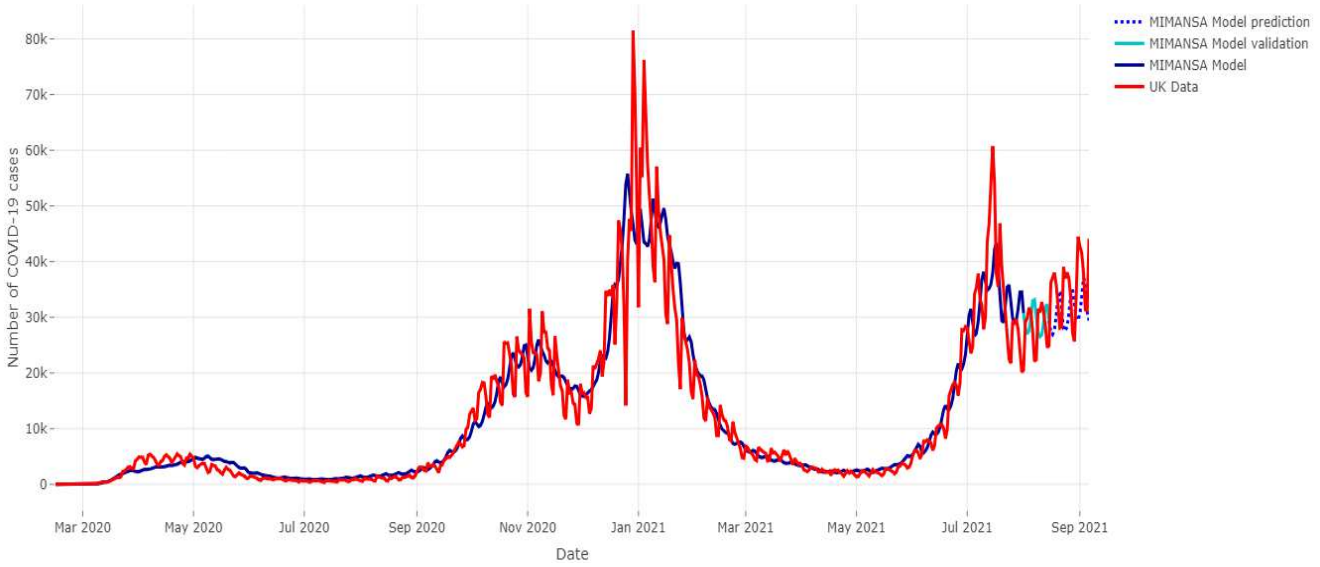



**(c)** MIMANSA Training, Validation, Prediction for California cases (Daily from July 1, 2021)

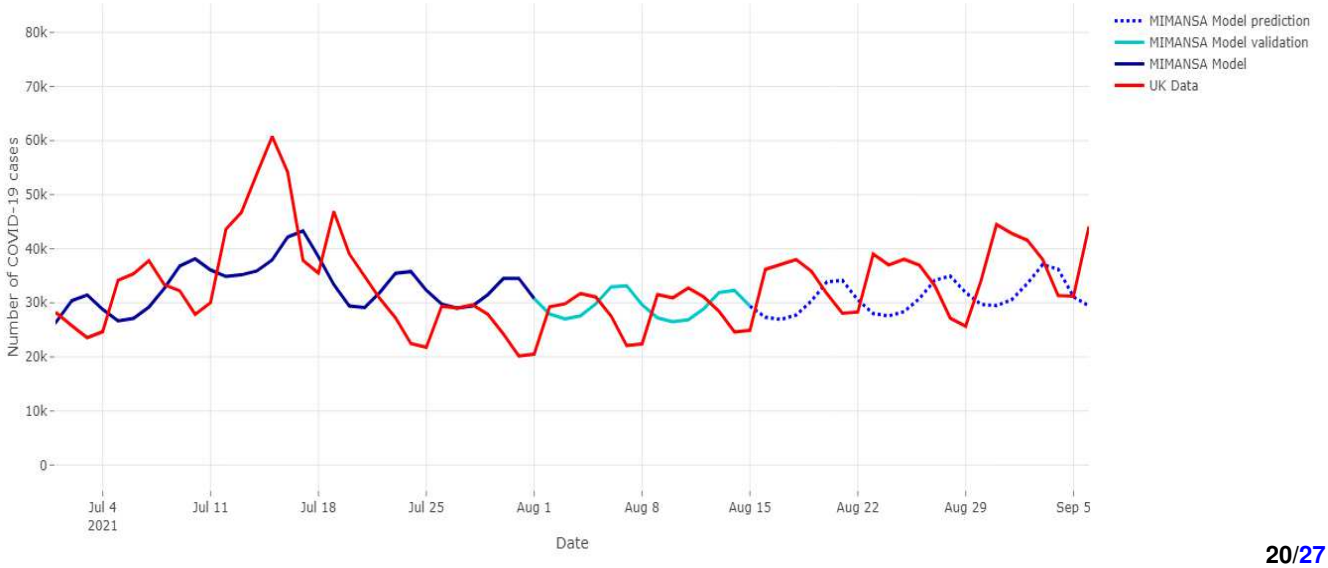
**Figure 8** California




**(a)** MIMANSA Training, Validation, Prediction for United Kingdom cases (Cumulative)




**(b)** MIMANSA Training, Validation, Prediction for United Kingdom cases (Daily)






**(c)** MIMANSA Training, Validation, Prediction for United Kingdom cases (Daily from July 1, 2021)

**Figure 9** United Kingdom

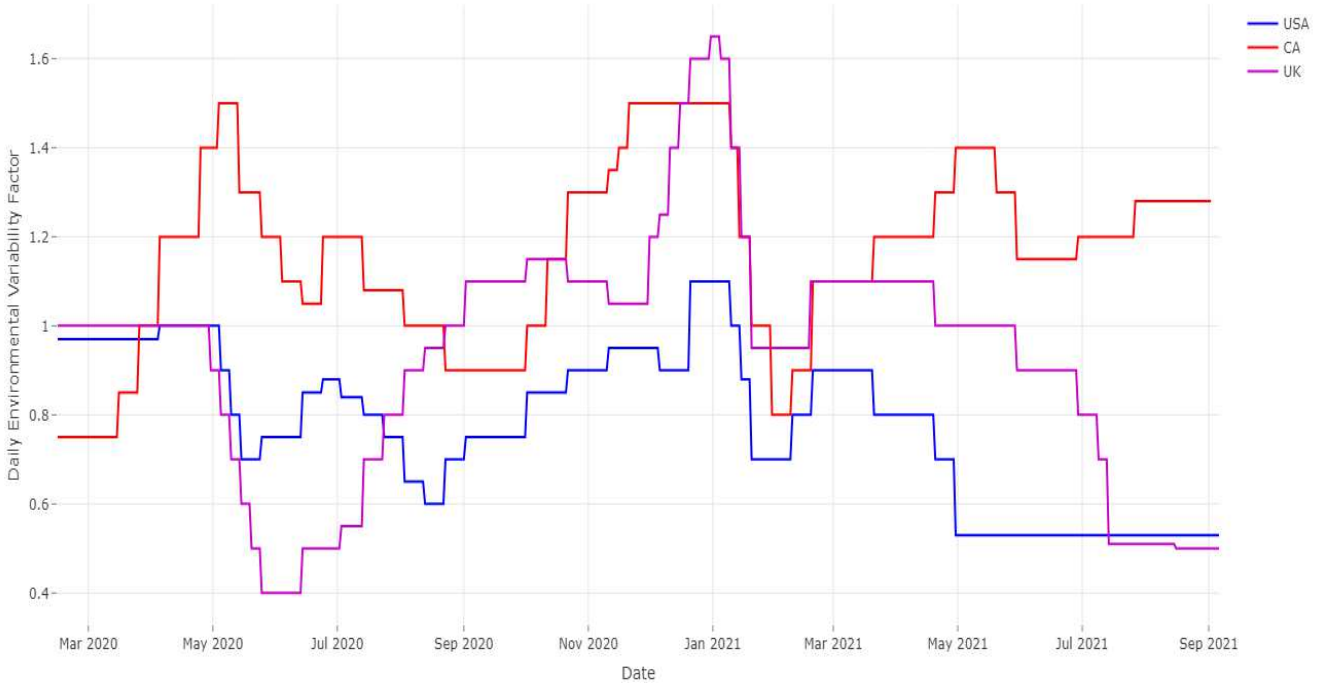



**Figure 10** Daily Environmental Variability

***Predictive capability***

Often it is asked how far in time can we project the number of COVID-19 cases. Our model depends on four major factors, lockdown, mask usage, presence of variants, and the percentage of vaccinated people. Predicting all 4 factors over a long duration is not easy due to multiple reasons. Lockdown and mask usage are policy-dependent as well as dependent on social behavior. Guessing when a new variant with a different transmission rate would come is not possible with the current knowledge.

***Data requirements***

MIMANSA can be trained on any regional dataset. The model needs the efficacy of the vaccine in percentage. Additionally, it needs the time series data for the following.

1. Mobility,
2. Mask usage,
3. variant growth in percentage, and
4. the percentage of vaccinated people.

In case any of the required data series is not available, one can make a reasonable guess. However, under those conditions, predictive performance may suffer.

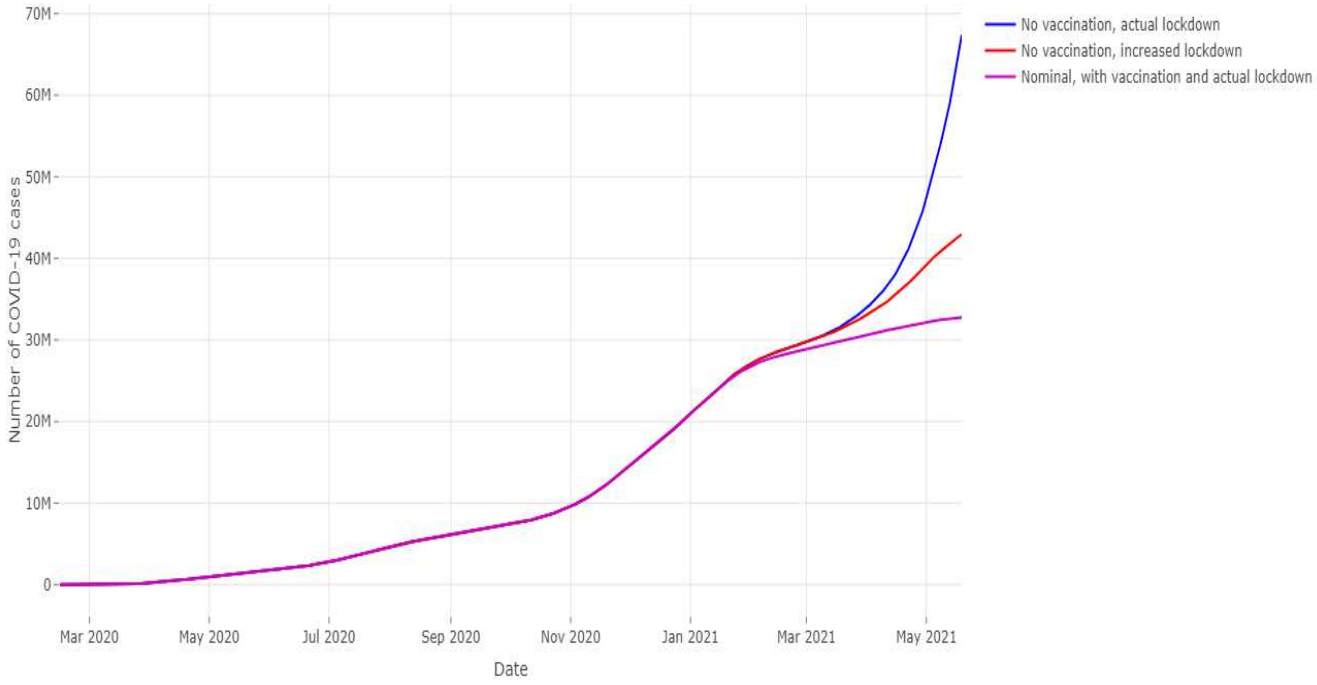
The data for the number of COVID-19 cases in the US was used to train the model. During the training phase, the data accurately fit the proposed model for over 9 months.

MIMANSA's novelty lies in its ability to represent the reality of COVID-19 spread and includes most of the observations in one system to give a complete overview of the dynamics of spread.

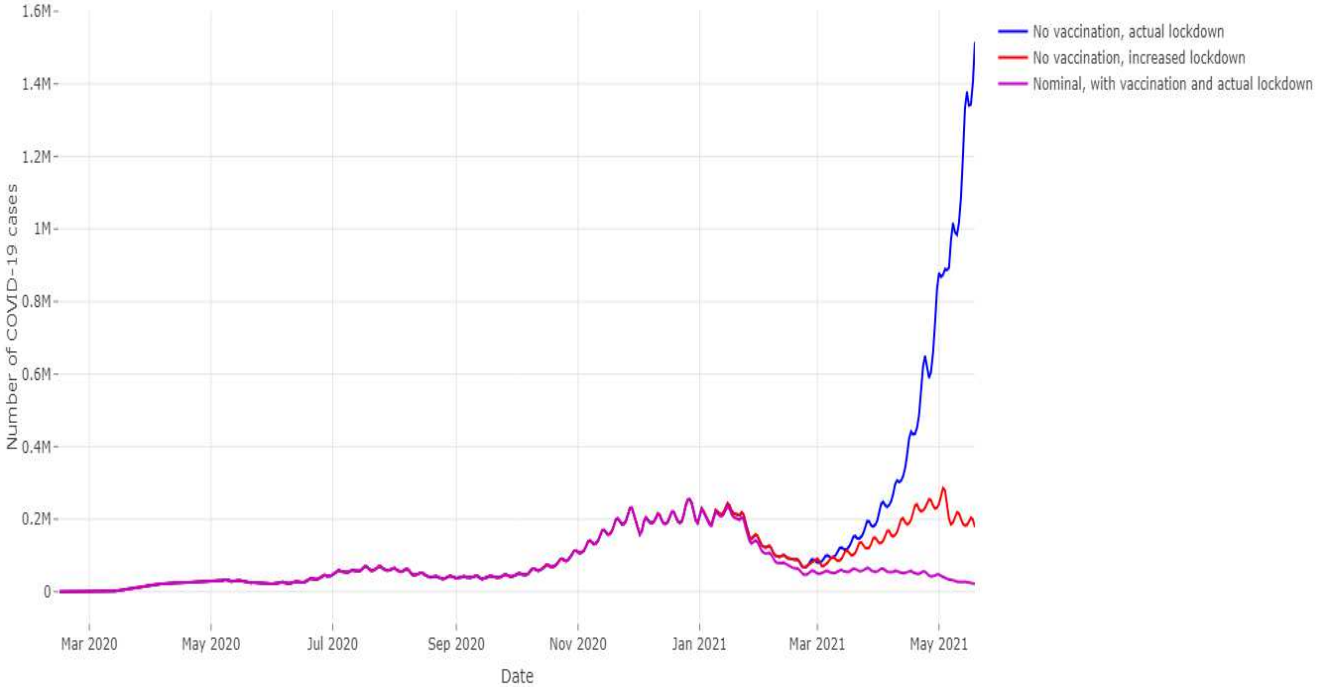
This model helps one simulate scenarios to study the impact of many different conditions. It handles sudden outbursts of cases like the spread of COVID-19 cases due to any event of mass gathering. It assists public health officials in complex decision-making, enables scientists in projecting the SARS-CoV-2 virus spread, and aids hospital administrators in the management of COVID-19 patients better.

MIMANSA is not a COVID-19 specific model. It is a generic model that is capable of simulating any virus spread. For a different virus, one has to change the exposure rates, the applied probability distributions, the incubation period, the infection rate, and perhaps some assumptions to reflect the characteristics of the new virus.

MIMANSA model is based on the current understanding of the disease. In the future, advancements in three areas that would improve our understanding are the effect of transmissibility of various strains of SARS-CoV-2, vaccinated population,

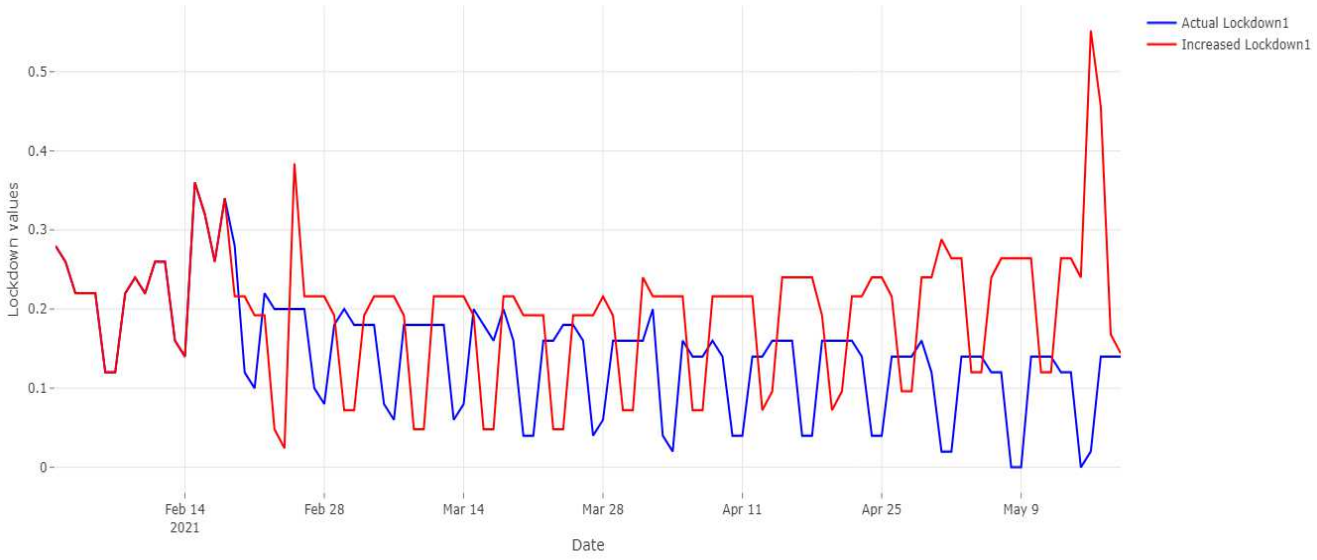



**(a)** Effect of Vaccination (Cumulative)

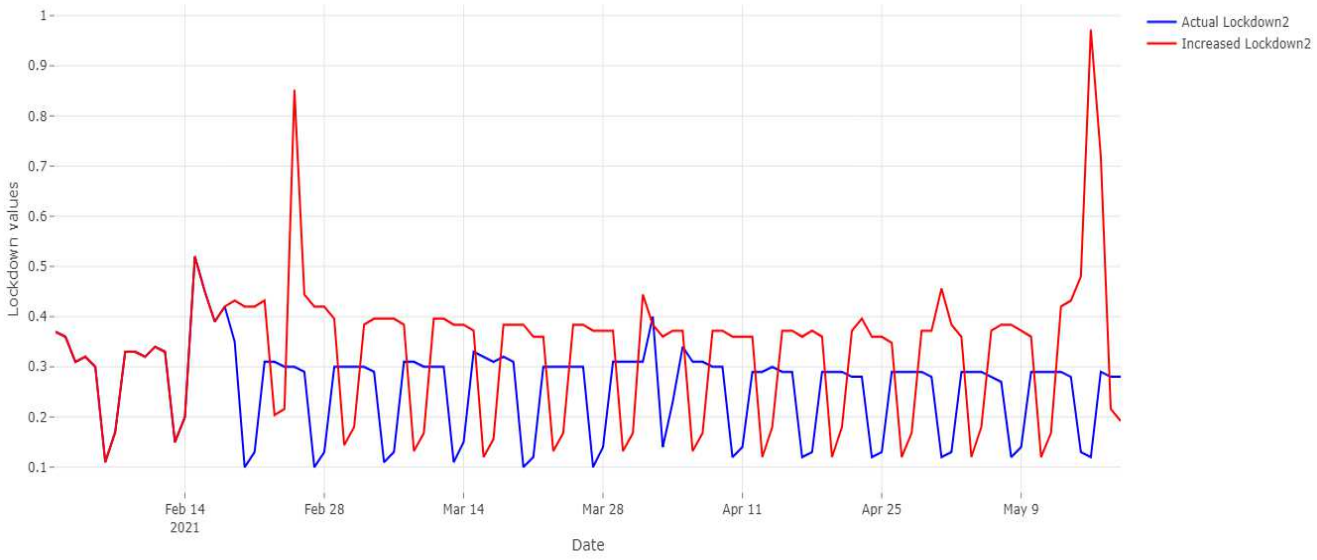



**(b)** Effect of Vaccination (Daily)

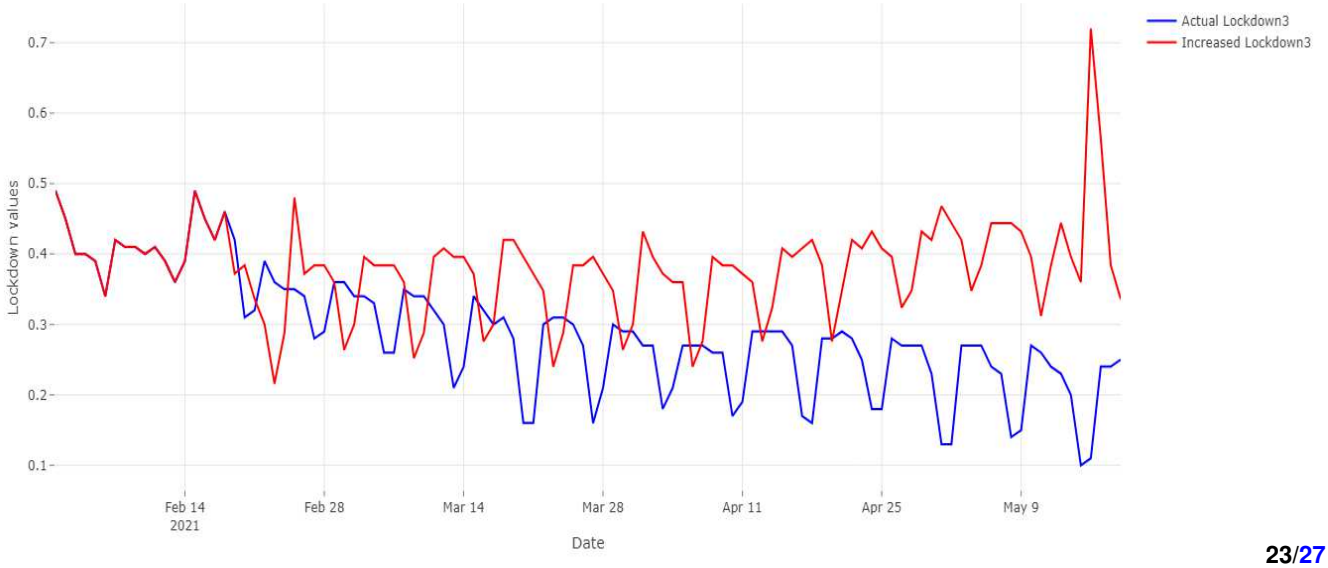
**Figure 11** Effect of Vaccination




**(a)** Lockdown1




**(b)** Lockdown2






**(c)** Lockdown3

**Figure 12** Actual lockdown and increased lockdown values used to see the effect of vaccination

and understanding of herd immunity. These factors will undoubtedly change the course of the spread of the virus. Incorporating their impact on the spread as multiple models would make MIMANSA more useful in the long run.

## Conclusion

The primary objective of any good COVID-19 model is to be able to evaluate the impact of any possible changes or perturbations in the ecosystem. For a good model, it should be able to consider regional changes in the distribution of variants, the transmissibility of each variant, the vaccinated population percentage, the mask usage, as well as mobility. Additionally, the model should take into account the regional environmental factors and the daily environmental variability factors. In this paper, we have proposed a model called MIMANSA and it satisfies all the above-listed requirements.

MIMANSA, the model for COVID-19, comprises representations of major observations, research findings, clinical findings, and every individual's social interaction into a mathematical form. The multiple representations are tied together logically to show how the SARS-CoV-2 spreads. The virus builds every day as a layer, connects each day by the next day in the form of a new layer. These layers form the virus proliferation network.

MIMANSA has multiple models inbuilt. We conducted an extensive study to find out which distribution represents the incubation period the best. Our study showed that the Weibull distribution is the best. Thus, we use the Weibull for simulating the incubation period inside MIMANSA. It has models for the exposure rate, silent carrier rate, and infection rate. All of these equations use the regional environmental variability as well as the daily environmental variability as parameters.

The model MIMANSA is built by keeping the right balance of model complexity vs. model simplicity that loses the elegance and fine details. We have reduced the number of parameters in the model by replacing variables with parameters by extensive literature search. With this, we were able to bring the number of training parameters down to 5.

This paper covers an in-depth study of COVID-19 growth in the number of cases in the USA, the UK, and California. We used actual data from these 3 places and trained the MIMANSA model. Further, we validated the model for each region and made predictions.

Our results show that in the number of cumulative cases the MIMANSA model predicted within an error margin of 2%.

In the end, we have also simulated a hypothetical scenario. Under this scenario, we presumed that no one in the USA was vaccinated till the end of June 2021. This analytical study clearly shows how important the vaccines have been in controlling the growth of SARS-CoV-2, thereby controlling the COVID-19 cases.

The objective of MIMANSA is not only to predict the number of people infected with the virus but also to provide a sound understanding of the factors responsible for the spread of any virus and provide the holistic view of virus spread that help make complex decisions to control virus spread. Additionally, it assists in understanding secondary attack rate, and trend analysis of COVID-19 cases. Moreover, the strength of the model is that it enables one to study the effects of lockdown, mask usage, household quarantine, vaccination, and transmissibility of variants of SARS-CoV-2 all at the same time making it a useful tool for policymakers and researchers.

## References


**1.** Medina-Mendieta, J. F., Cortés-Cortés, M., & Cortés-Iglesias, M. COVID-19 Forecasts for Cuba Using Logistic Regression and Gompertz Curves. *MEDICC review*. **22** (2020).

**2.** Maleki, M., Mahmoudi, M. R., Heydari, M. H., & Pho, K. H. Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models. *Chaos, Solitons & Fractals*. **140**, 110151 (2020).

**3.** Malki, Z., et al. ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. *Neural Computing and Applications*. **33**, 2929-2948 (2021).

**4.** Khakharia, A., et al. Outbreak prediction of COVID-19 for dense and populated countries using machine learning. *Annals of Data Science*. **8**, 1-19 (2021).

**5.** Kucharski, A. J. et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*. **20**, 553-558 (2020).

**6.** Grassly, N. & Fraser, C. Mathematical models of infectious disease transmission. *Nat Rev Microbiol*. **6**, 477–487 (2008).

**7.** Klinkenberg, D., Fraser, C. & Heesterbeek, H. The effectiveness of contact tracing in emerging epidemics. *PloS one*. **1**, (2006).

**8.** Mandal, S. et al. Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach. *Indian J. Med. Res*. **151**, 190-199 (2020).

**9.** Fanelli, D & Piazza, F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Soliton Fract*. **134**, 109761 (2020).

**10.** Moein, S. et al. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. *Scientific Reports*. **11**, 1-9 (2021).

**11.** Von Bertalanffy & L. *General System Theory: Foundations, Development, Applications*. (1968).

**12.** Vaidya, V. G., Alexandro Jr, Frank. J. & Vaidya, P. G. A bilinear control model for tumor growth and chemotherapy. *Int J Biomed Comput*. **17**, 145-154 (1985).

**13.** Vaidya, P. G., Vaidya, V. G. & Martin D. G. An application of the non-linear bifurcation theory to tumor growth modeling. *Int J Biomed Comput*. **27**, 27-46 (1991).

**14.** JAMA network website, https://jamanetwork.com/journals/jama/article-abstract/2762028.

**15.** Bai, Y. et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA. **323**, 1406–1407 (2020).

**16.** Hu, Z. et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci*. **63**, 706-711 (2020).

**17.** Lauer, S. A. et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med*. **172**, 577-82 (2020).

**18.** Daily website, https://www.sciencedaily.com/releases/2020/03/200310164744.htm.

**19.** National Institutes of Health website, https://www.nih.gov/news-events/news-releases/new-coronavirus-stable-hours-surfaces/.

**20.** Del Valle, S. Y., Hyman, J.M., Hethcote, H. W. & Eubank S.G. Mixing patterns between age groups in social networks. *Soc. Netw.* **29**, 539-54 (2007).

**21.** Dugdale, C. M., & Walensky, R. P. Filtration efficiency, effectiveness, and availability of N95 face masks for COVID-19 prevention. *JAMA Internal Medicine* **180**, 1612-1613 (2020).

**22.** MIT Medical website, https://medical.mit.edu/covid-19-updates/2020/07/how-long-symptom-onset-person-contagious.

**23.** Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J. & Hsueh, P. R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents*. **55**, 105924 (2020).

**24.** Patel, A. P., Welling, A. A., Vaidya, V. G., & Kulkarni, P. S. Simulation of COVID-19 Incubation Period and the Effect of Probability Distribution Function on Model Training Using MIMANSA. *medRxiv* doi:https://doi.org/10.1101/2020.06.18.20134460. (2020).

**25.** Quora website, https://www.quora.com/search?q=transmissibility%20of%20SARS%20-cov-2%20variant&author=120885241

**26.** The Conversation website, https://theconversation.com/1-in-4-unvaccinated-people-may-not-comply-with-cdc-guidelines-to-wear-masks-indoors-survey-suggests-161336.

**27.** AXIOS website, https://www.axios.com/ipsos-poll-mask-use-vaccinated-americans-66422391-8496-4035-88fa-70312a2b288d.html.

**28.** OWID website, https://ourworldindata.org/covid-cases.

**29.** LA Times website, https://github.com/datadesk/california-coronavirus-data/blob/master/latimes-state-totals.csv#L1.

**30.** GOV.UK website, https://coronavirus.data.gov.uk/details/cases.

**31.** CMR website, https://www.google.com/covid19/mobility/.

**32.** IHME website, http://www.healthdata.org/acting-data/maps-mask-use/.

**33.** BCC website, https://www.bbc.com/news/health-51205344.

**34.** Office for national statistics, https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/articles/twothirdsofadultsstillplantowearmasksinshopsandonpublictransport/2021-07-16.

**35.** Wang, Y. et al. Transmission, viral kinetics and clinical characteristics of the emergent SARS-CoV-2 Delta VOC in Guangzhou, China. *EClinicalMedicine* **40**, 101129. (2021).

**36.** Outbreak website, https://outbreak.info/situation-reports#voc.

**37.** BCC website, https://www.bbc.com/future/article/20210114-covid-19-how-effective-is-a-single-vaccine-dose.

**38.** USAFacts website, https://usafacts.org/visualizations/covid-vaccine-tracker-states/state/california.

**39.** USAFacts website, https://static.usafacts.org/public/data/external/COVID19_CDC_Vaccination_CSV_Download.csv.

**40.** Yanes-Lane, M. et al. Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis. *PloS one*. **15**, e0241536. https://doi.org/10.1371/journal.pone.0241536. (2020).

**41.** Qiu, X. et al. Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission–a living systematic review. Clinical microbiology and infection. Preprint at https://doi.org/10.1101/2020.09.01.20135194 (2021).

**42.** Yang, Q., et al. Just 2% of SARS-CoV-2 positive individuals carry 90% of the virus circulating in communities/ *Proceedings of the National Academy of Sciences*. **118** (2021).

## Acknowledgments

We thank Dr. Abhishek Karwa, Asst. Professor of Medicine, School of Medicine, University of California San Francisco, California, for meticulously going through multiple drafts of the paper and making valuable suggestions. His inputs have been instrumental in shaping this paper.

We thank Dr. Victor Grubsky, Director of Engineering at Mercury Systems, Torrance, California, for pointing out gaps in the paper and asking thought-provoking questions. His guidance helped us improve the manuscript and bring more clarity in writing. As a result, the quality of the paper has been enhanced.

## Author contributions statement

AW and AP contributed equally. Their primary contribution was in the conceptualization, design, software implementation, and complete data analysis. PK was the medical advisor. He ensured a medically accurate system design. VV was the principal investigator. His contribution was in the areas of conceptualization, mathematical formulation of the known medical facts, system design, simulation, and data analysis. All authors contributed equally in reviewing the manuscript.

## Additional information

## Compliance with ethical standards

## Conflict of interest

The authors declare that they have no conflict of interest.

## Ethics approval

The research is based on data available over the internet. Thus, no ethics approvals are required.

## Consent to participate

There was no participation of individuals in this study.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Consent to publish

No human participation was involved in this study.

## Availability of data and material

The data set is available on the internet and the sources have been listed. Additionally, the dataset used will be made available upon request.

## Code availability

The codes used will be made available upon request. Please contact the corresponding author.

## Appendix

Parameter Set, $S[X_i]$, for the USA, CA and the UK used in the equation (38),

United States of America, USA:
S[ER1] = [0.9705, 0.1, 3.4,0.265,0,0]
S[ER2] = [0.64, 0.1, 3.4, -1.9,2, -0.075]
S[ER3] = [0.5988, 0.1, 3.4, -1.9,2, -0.075]
S[SCR] = [0.7,0,0, -1.9,2, -0.075]
S[IR] = [0.48,0,0, -1.9,2, -0.075]

California, CA:
S[ER1] = [0.9553, 0.1, 3.4,0.265,0,0]
S[ER2] = [0.63, 0.1, 3.4, -1.9,2, -0.075]
S[ER3] = [0.5894, 0.1, 3.4, -1.9,2, -0.075]
S[SCR] = [0.7,0,0, -1.9,2, -0.075]
S[IR] = [0.48,0,0, -1.9,2, -0.075]

United Kingdom, UK:
S[ER1] = [0.7581, 0.1, 3.4,0.265,0,0]
S[ER2] = [0.5, 0.1, 3.4, -1.9,2, -0.075]
S[ER3] = [0.4677, 0.1, 3.4, -1.9,2, -0.075]
S[SCR] = [0.62,0,0, -1.9,2, -0.075]
S[IR] = [0.48,0,0, -1.9,2, -0.075]