

# Estimation of Time-Dependent Transmission Rate for COVID-19 SVIRD Model Using Predictor–Corrector Algorithm



Ruiyan Luo, Alejandra D. Herrera-Reyes, Yena Kim, Susan Rogowski, Diana White, and Alexandra Smirnova

## 1 Introduction

Compartmental disease models, which track the progression of individuals between different disease stages and risk levels, remain at the kernel of epidemic theory [1]. A simple example of a compartmental framework is the Susceptible–Infected–Recovered (SIR) model proposed in [2]. This model has been extended to include other states, such as the Susceptible–Infectious–Recovered–Deceased (SIRD) [3] and the Susceptible–Infectious–Recovered–Vaccinated (SIRV) models

---

R. Luo

Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA

e-mail: [rluo@gsu.edu](mailto:rluo@gsu.edu)

A. D. Herrera-Reyes

Department of Mathematical Science, George Mason University, Fairfax, VA, USA

e-mail: [aherre6@gmu.edu](mailto:aherre6@gmu.edu)

Y. Kim

Hawaii Pacific University, Honolulu, HI, USA

e-mail: [yekim@hpu.edu](mailto:yekim@hpu.edu)

S. Rogowski

Department of Mathematics, Florida State University, Tallahassee, FL, USA

e-mail: [srogowski@fsu.edu](mailto:srogowski@fsu.edu)

D. White

Department of Mathematics & Statistics, Clarkson University, Potsdam, NY, USA

e-mail: [dtwhite@clarkson.edu](mailto:dtwhite@clarkson.edu)

A. Smirnova (✉)

Department of Mathematics & Statistics, Georgia State University, Atlanta, GA, USA

e-mail: [asmirnova@gsu.edu](mailto:asmirnova@gsu.edu)

[4]. Recently, generalizations of SIR models have been implemented to study the spread of COVID-19 with the adherence and non-adherence of social behavior protocols such as masking, social distancing, and the enforcement of closures and lockdowns [5–9]. Earlier models described the spread of the disease in uncontrolled systems and in the presence of different mitigation strategies such as social distancing and lockdown restrictions.

Since the development and widespread distribution of vaccines, incorporation of vaccination into such models has been an important development [10, 11]. However, few models have accounted for differing disease transmission within vaccinated and unvaccinated individuals. Here, we propose a new compartmental model of COVID-19 transmission that takes into consideration some of these important dynamics by including the vaccination status of both susceptible and infected humans. We also include the possibility of losing immunity and becoming reinfected within both vaccinated and unvaccinated populations. Thus, our new model incorporates important disease dynamics that have not been covered by previous COVID-19 models. Additionally, the proposed model can easily be adjusted to other seasonal outbreaks. With new variants of COVID-19 and other viruses occurring regularly, along with fluctuations of vaccine efficacy among these variants, this new model will help to understand past and current disease dynamics and make predictions about future cases.

Another important novel feature of our compartmental model is the use of a time-dependent transmission rate. Oftentimes, the transmission rate of a disease is the most challenging parameter to estimate [12]. The emerging new variants of COVID-19 make stable estimation of disease transmission even more complicated. To simplify this, many previous COVID-19 models incorporated constant transmission rates found in the literature. To better assess the efficiency of control and prevention and to account for new COVID-19 strains, in our proposed model, we introduce a time-dependent transmission rate for vaccinated and unvaccinated individuals. This rate is reconstructed from noise-contaminated data on new incidence cases and daily deaths by solving a parameter estimation inverse problem.

A commonly used method for estimating parameters of ordinary differential equations (ODEs) from noisy data is nonlinear least squares (NLS), where model predictions for an invading pathogen are fitted to reported incidence cases and daily new deaths [13–16]. In the NLS, a numerical method, such as Runge–Kutta or similar, is used to approximate the solution of a given ODE system using a trial set of values for parameters and initial conditions. The fit value is then input into an optimization algorithm that updates parameter estimates. As a result, the NLS framework can be computationally expensive when noisy data is considered or a highly nonlinear model is being used to describe a complex biological process. In [17, 18], a two-stage approach for this method was proposed, which first fit a smooth curve to given noisy data and then estimated the unknown parameters in the ODE system. Ramsay et al. [19] expanded on this method by proposing to alternate the two procedures and by imposing a smoothness penalty on curve fitting. To that end, Ramsay et al. developed a novel profiling estimation procedure where the data fitting and the fidelity to the ODE were combined into a penalized log-likelihood criterion,

which provided the statistical inference for the ODE parameters. For other prior work on alternating minimization, also known as (block) coordinate descent, one may consult [20–25] and the references therein.

A more general nonlinear constrained minimization problem was studied in [26], where parameter estimation was carried out in a predictor–corrector manner. In the predictor–corrector algorithm of [26], one updates the epidemiological parameters by a regularized second-order method while freezing the state variables, and then the state variables are modified while the system (epidemiological) parameters are fixed. These updates are iterated until convergence. Here, we propose a new predictor–corrector algorithm that extends the earlier version in [26] to the case of parameter-dependent nonlinear observation operators. The new algorithm successfully mitigates the associated computational costs and incorporates an extra layer of stability in the optimization process. In what follows, the proposed version of the predictor–corrector algorithm is used to get stable estimates of a time-dependent transmission rate and effective reproduction number from our new compartmental model, which is applied to the study of COVID-19 dynamics in a post-vaccination stage.

The chapter is organized as follows. In Sect. 2, we introduce our Susceptible–Vaccinated–Infectious–Recovered–Deceased (SVIRD) model. In Sect. 3, we describe the new computational algorithm for estimating disease parameters in the proposed epidemic model. In Sects. 4 and 5, the method is evaluated on synthetic and real data sets, respectively. Possible directions of future work are outlined in Sect. 6.

## 2 Mathematical Model: SVIRD

Prior studies have underscored the importance of stable parameter estimation related to infectious disease transmission models based on ordinary or partial differential equations [27–29]. Lack of stable parameter estimation, which is evident when parameter estimates are associated with large uncertainties, may be attributed to the model structure or to the lack of information in a given data set, which could be linked to the number of observations and to the spatial granularity of the data [28].

Within epidemiology, stable estimation of the effective reproduction number,  $\mathcal{R}_e(t)$ , and its underlying transmission rate,  $\beta(t)$ , is particularly important [30–32]. Unlike other system parameters, i.e., incubation and recovery rates, the effective reproduction number and the transmission rate of the disease are directly influenced by mitigation measures. Therefore, it is critical to develop both suitable epidemic models and regularized computational methods to reliably quantify disease-specific parameters, especially in the face of noise-contaminated data and substantial uncertainty in approximate solutions.

In this chapter, to model the COVID-19 dynamics and estimate the effective reproduction number,  $\mathcal{R}_e(t)$ , and its underlying transmission rate,  $\beta(t)$ , we propose the following system of ODEs:

$$\frac{dS}{dt} = -\beta(t) \frac{S(t)}{N - D(t)} (I_s(t) + I_v(t)) - pS(t) + \delta_r R(t) + \delta_v V(t) \quad (1)$$

$$\frac{dV}{dt} = pS(t) - (1 - \alpha)\beta(t) \frac{V(t)}{N - D(t)} (I_s(t) + I_v(t)) - \delta_v V(t) \quad (2)$$

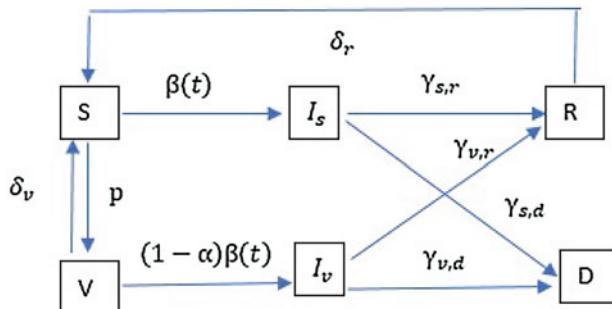
$$\frac{dI_s}{dt} = \beta(t) \frac{S(t)}{N - D(t)} (I_s(t) + I_v(t)) - (\gamma_{s,r} + \gamma_{s,d}) I_s(t) \quad (3)$$

$$\frac{dI_v}{dt} = (1 - \alpha)\beta(t) \frac{V(t)}{N - D(t)} (I_s(t) + I_v(t)) - (\gamma_{v,r} + \gamma_{v,d}) I_v(t) \quad (4)$$

$$\frac{dR}{dt} = \gamma_{s,r} I_s(t) + \gamma_{v,r} I_v(t) - \delta_r R(t) \quad (5)$$

$$\frac{dD}{dt} = \gamma_{s,d} I_s(t) + \gamma_{v,d} I_v(t). \quad (6)$$

The system defined by Eqs. (1)–(6) includes susceptible unvaccinated ( $S$ ), susceptible vaccinated ( $V$ ), infected vaccinated ( $I_v$ ), infected unvaccinated ( $I_s$ ), recovered ( $R$ ), and deceased ( $D$ ) compartments. With  $N$  denoting the population size at the beginning time point of the study period, we use  $N - D(t)$  as the total population size at time  $t$ . This is based on the assumption that the population increase (due to birth or immigration) and population decrease (due to reasons other than COVID-19) balance out, and the change in population size is just due to COVID-19 death. The diagram of the SVIRD model in Eqs. (1)–(6) is given in Fig. 1, which illustrates the transition of individuals between various disease compartments. Susceptible humans become vaccinated at a rate of  $p$ . Both vaccinated and unvaccinated individuals can be infected. The disease transmission rate,  $\beta(t)$ , for susceptible



**Fig. 1** Diagram of the SVIRD model used. Susceptible individuals get vaccinated at a rate  $p$  and become infected at a time-dependent transmission rate  $\beta(t)$ . A constant parameter,  $0 < \alpha < 1$ , is a measure of vaccine efficacy. The lower values correspond to less efficacy, and  $(1 - \alpha)\beta(t)$  is the rate of disease transmission for vaccinated individuals. Both infected unvaccinated and vaccinated can recover at rates  $\gamma_{s,r}$  and  $\gamma_{v,r}$  and die at rates  $\gamma_{s,d}$  and  $\gamma_{v,d}$ , respectively. Loss of immunity is accounted for by considering movement back to the susceptible class from the vaccinated and recovered classes at rates  $\delta_v$  and  $\delta_r$

individuals is assumed to be time-dependent. We assume that vaccinated individuals become infected at a slower rate, which is taken into account by the incorporation of a vaccine efficacy parameter, denoted by  $\alpha$ ; that is, vaccinated individuals become infected at a rate of  $(1 - \alpha)\beta(t)$ , where  $0 < \alpha < 1$ .

Motivated by the report that unvaccinated individuals are more likely to have severe symptoms from COVID-19 infections leading to a higher risk of hospitalization and death [33], we assume different death rates for vaccinated and unvaccinated individuals, denoted by  $\gamma_{v,d}$  and  $\gamma_{s,d}$ , respectively. The severity in symptoms also leads to differing recovery rates for vaccinated and unvaccinated populations. The recovery rates for vaccinated and unvaccinated individuals are denoted by  $\gamma_{v,r}$  and  $\gamma_{s,r}$ , respectively.

We further consider the case of possible reinfection due to the loss of immunity by vaccinated individuals at a rate of  $\delta_v$  and recovered individuals at a rate of  $\delta_r$ . We note from Eq. (1) that the rate of transmission depends only on the number of contacts between the living susceptible and infected individuals (described by the division by  $N - D(t)$ , the total living population at any instance in time).

The disease transmission rate,  $\beta(t)$ , is an important underlying factor for the effective reproduction number,  $\mathcal{R}_e(t)$ , which quantifies the number of secondary cases per primary case in a completely susceptible population during the entire course of the outbreak. Similar to the transmission rate, the effective reproduction number is significantly impacted by environmental conditions and the behavior of the population. A sustainable reduction of  $\mathcal{R}_e(t)$  to a level less than 1 would indicate that mitigation measures are successful and that the disease is contained, because every infected person, on average, can only transmit the virus to less than one other human.

Using the next-generation matrix [34, 35], the effective reproduction number for compartmental model (Eqs. (1)–(6)) is estimated as

$$\mathcal{R}_e(t) = \frac{\beta(t)}{(\gamma_{s,r} + \gamma_{s,d})} \frac{S(t)}{N - D(t)} + \frac{(1 - \alpha)\beta(t)}{(\gamma_{v,r} + \gamma_{v,d})} \frac{V(t)}{N - D(t)}. \quad (7)$$

From Eq. (7), we note that  $\mathcal{R}_e(t)$  increases with increasing disease transmission  $\beta(t)$ , as well as increasing numbers of susceptible individuals (vaccinated and unvaccinated). In addition,  $\mathcal{R}_e(t)$  decreases with increasing recovery rates. Next, in Sect. 3, we describe our predictor–corrector algorithm that will be used to reconstruct the disease transmission rate,  $\beta(t)$ , which allows us to provide an estimate for the effective reproduction number,  $\mathcal{R}_e(t)$ .

### 3 Methodology and Algorithm

Let  $C$  and  $\mathcal{T}$  be incidence data on new COVID-19 confirmed cases and deaths, respectively, and  $n$  be the number of data points in each set. Naturally, we assume that both data sets are noise contaminated. According to our SVIRD model given

by Eqs. (1)–(6), the daily number of new COVID-19 cases is

$$\mathbb{C}(t) := \beta(t) \frac{S(t)(I_s(t) + I_v(t))}{N - D(t)} + (1 - \alpha)\beta(t) \frac{V(t)(I_s(t) + I_v(t))}{N - D(t)}, \quad (8)$$

which we define as the rate of new infections into the system. On the other hand, by Eq. (6), the daily number of new deaths is

$$\mathbb{T}(t) := \gamma_{s,d} I_s(t) + \gamma_{v,d} I_v(t). \quad (9)$$

Assume that in a particular region, the values  $a = t_1$  and  $b = t_n$  are the first and the last days of the study period. We note that, fortunately, the number of deceased individuals is considerably smaller than infectious ones. So, we multiply daily new deaths,  $\mathbb{T}$ , by a positive scaling parameter,  $\lambda$ , to ensure that new deaths and new cases have the same order of magnitude. Let the data,  $d$ , for new cases and deaths,  $C$  and  $\mathcal{T}$ , be reported on days  $t_1, t_2, \dots, t_n$ . That is,

$$d := [C(t_1), \dots, C(t_n), \lambda\mathcal{T}(t_1), \dots, \lambda\mathcal{T}(t_n)]^T. \quad (10)$$

Combining Eqs. (8) and (9), we now introduce the observation operator as

$$\mathcal{B} := [\mathbb{C}(t_1), \dots, \mathbb{C}(t_n), \lambda\mathbb{T}(t_1), \dots, \lambda\mathbb{T}(t_n)]^T. \quad (11)$$

Then our goal is to recover the unknown time-dependent transmission rate,  $\beta(t)$ , from the nonlinear constrained minimization problem:

$$\min_{\beta, S, V, I_s, I_v, D} f(\beta, S, V, I_s, I_v, D) \quad (12)$$

subject to system in Eqs. (1)–(6), where

$$\begin{aligned} f(\beta, S, V, I_s, I_v, D) := & \|\mathcal{B} - d\|^2 \\ & = \sum_{i=1}^n \left\{ (\mathbb{C}(t_i) - C(t_i))^2 + \lambda^2 (\mathbb{T}(t_i) - \mathcal{T}(t_i))^2 \right\}. \end{aligned} \quad (13)$$

To solve Eqs. (12) and (13) numerically, we discretize unobserved state variables,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , and the time-varying transmission rate,  $\beta(t)$ , using basis expansions. The vector of expansion coefficients for the transmission rate,  $\beta(t)$ , is of primary interest. The vector of expansion coefficients for the state variables is of less practical importance, and it is primarily needed for the estimation of  $\beta(t)$ . For this reason, in statistics literature, the expansion coefficients for state variables are often referred to as nuisance parameters [19]. Upon discretization, we iteratively update both sets of unknown expansion coefficients using alternating minimization as described below.

In order to obtain the discrete approximation of  $\beta(t)$ , we consider a finite subset spanned by shifted Legendre polynomials of degree 0, 1, ...,  $m - 1$ , which are orthogonal on the interval  $[a, b]$  with respect to  $L_2$  inner product, defined recursively as follows:

$$x = \frac{2t - a - b}{b - a}, \quad P_0(x) = 1, \quad P_1(x) = x, \quad t \in [a, b],$$

$$(j + 1)P_{j+1}(x) = (2j + 1)xP_j(x) - jP_{j-1}(x), \quad j = 1, 2, \dots, m - 2.$$

This gives rise to the following finite-dimensional approximation of the transmission rate:

$$\bar{\beta}_i[\theta] = \sum_{j=0}^{m-1} \theta_{j+1} P_j(t_i), \quad i = 1, 2, \dots, n. \quad (14)$$

Likewise, we express the state variables  $S$ ,  $V$ ,  $I_s$ , and  $I_v$  as

$$\begin{aligned} \bar{S}_i[u] &= \sum_{j=0}^{l-1} u_{j+1} P_j(t_i), & \bar{V}_i[u] &= \sum_{j=0}^{l-1} u_{l+j+1} P_j(t_i), \\ \bar{I}_{s,i}[u] &= \sum_{j=0}^{l-1} u_{2l+j+1} P_j(t_i), & \bar{I}_{v,i}[u] &= \sum_{j=0}^{l-1} u_{3l+j+1} P_j(t_i), \end{aligned} \quad (15)$$

which generates discretized daily rates of incidence and death,  $\bar{C}_{d,i}[\theta, u]$  and  $\bar{T}_{d,i}[u]$ , respectively, if one substitutes  $\bar{\beta}_i[\theta]$  from Eq. (14) and  $\bar{S}_i[u]$ ,  $\bar{V}_i[u]$ ,  $\bar{I}_{s,i}[u]$ , and  $\bar{I}_{v,i}[u]$  from Eq. (15) for  $\beta(t_i)$ ,  $S(t_i)$ ,  $V(t_i)$ ,  $I_s(t_i)$ , and  $I_v(t_i)$  in Eqs. (1)–(6) and Eqs. (8) and (9). The derivatives of  $S$ ,  $V$ ,  $I_s$ , and  $I_v$  get discretized by replacing  $P_j(t_i)$  with  $P'_j(t_i)$  in the identities above.

Next, we define vectors for the unknown parameters,  $\theta$  and  $u$ , from the discrete approximation of the transmission rate,  $\beta(t_i)$ , in identity Eq. (14) and from the discrete approximation of the state variables,  $S(t_i)$ ,  $V(t_i)$ ,  $I_s(t_i)$ , and  $I_v(t_i)$ ,  $i = 1, 2, \dots, n$ , in Eq. (15) as

$$\theta := [\theta_1, \dots, \theta_m]^T \text{ and } u := [u_1, \dots, u_l, u_{l+1}, \dots, u_{2l}, u_{2l+1}, \dots, u_{3l}, u_{3l+1}, \dots, u_{4l}]^T.$$

This enables us to introduce the observation operator,  $B$ :

$$B(\theta, u) := \left[ \bar{C}_{d,1}[\theta, u], \dots, \bar{C}_{d,n}[\theta, u], \lambda \bar{T}_{d,1}[\theta, u], \dots, \lambda \bar{T}_{d,n}[\theta, u] \right]^T \quad (16)$$

and the operator  $G$  to account for the constraints

$$\begin{aligned}
G_i(\theta, u) &:= \bar{S}'_i[u] + \bar{\beta}_i[\theta] \frac{\bar{S}_i[u](\bar{I}_{s,i}[u] + \bar{I}_{v,i}[u])}{N - \bar{D}_i[u]} + p\bar{S}_i[u] - \delta_r \bar{R}_i[u] - \delta_v \bar{V}_i[u] \\
G_{n+i}(\theta, u) &:= \bar{V}'_i[u] - p\bar{S}_i[u] + (1 - \alpha)\bar{\beta}_i[\theta] \frac{\bar{V}_i[u](\bar{I}_{s,i}[u] + \bar{I}_{v,i}[u])}{N - \bar{D}_i[u]} + \delta_v \bar{V}_i[u] \\
G_{2n+i}(\theta, u) &:= \bar{I}'_{s,i}[u] - \bar{\beta}_i[\theta] \frac{\bar{S}_i[u](\bar{I}_{s,i}[u] + \bar{I}_{v,i}[u])}{N - \bar{D}_i[u]} + (\gamma_{s,r} + \gamma_{s,d})\bar{I}_{s,i}[u] \\
G_{3n+i}(\theta, u) &:= \bar{I}'_{v,i}[u] - (1 - \alpha)\bar{\beta}_i[\theta] \frac{\bar{V}_i[u](\bar{I}_{s,i}[u] + \bar{I}_{v,i}[u])}{N - \bar{D}_i[u]} + (\gamma_{v,r} + \gamma_{v,d})\bar{I}_{v,i}[u]
\end{aligned}$$

for  $i = 1, 2, \dots, n$ . Here  $\bar{D}_i[u]$  is the reported cumulative number of deaths on day  $t_i$  and

$$\bar{R}_i[u] := N - (\bar{S}_i[u] + \bar{V}_i[u] + \bar{I}_{s,i}[u] + \bar{I}_{v,i}[u] + \bar{D}_i[u]). \quad (17)$$

We can now recast the constrained minimization problem as follows:

$$\begin{aligned}
\text{minimize} \quad & \|B(\theta, u) - d\|^2 \quad \text{with respect to } \theta \text{ and } u \\
\text{subject to} \quad & G(\theta, u) = 0.
\end{aligned} \quad (18)$$

Note that the data-fitting operator,  $B$ , also depends on the input data,  $\bar{D}$ , the cumulative number of deceased individuals. However, the cumulative data, as opposed to daily number of cases and deaths on the right-hand side, are smooth, and the noise in cumulative data is consistent with discretization and modeling errors.

To reconstruct the transmission rate,  $\beta(t)$ , we employ a predictor–corrector algorithm, where one updates  $\theta$  while freezing  $u$ , and then  $u$  is modified while  $\theta$  is kept unchanged. The process is repeated until a desired tolerance level is achieved.

More specifically, given  $\begin{pmatrix} \theta_k \\ u_k \end{pmatrix}$ , one transitions from  $\theta_k$  to  $\theta_{k+1}$  by applying one step of the iteratively regularized Gauss–Newton (IRGN) procedure:

$$\begin{aligned}
\theta_{k+1} &= \theta_k - [G'_\theta(\theta_k, u_k)G'_\theta(\theta_k, u_k) + B'^*_\theta(\theta_k, u_k)B'_\theta(\theta_k, u_k) + \tau_k I]^{-1} \\
&\quad \{G'^*_\theta(\theta_k, u_k)G(\theta_k, u_k) + B'^*_\theta(\theta_k, u_k)(B(\theta_k, u_k) - d) + \tau_k(\theta_k - \bar{\theta})\}, \quad (19)
\end{aligned}$$

where  $\tau_k$  is the regularization parameter needed to incorporate stability in the optimization process and  $\bar{\theta}$  is a prior value of  $\theta$ . Then, given  $\begin{pmatrix} \theta_{k+1} \\ u_k \end{pmatrix}$ , one computes  $u_{k+1}$  using the classical Gauss–Newton scheme

$$\begin{aligned}
u_{k+1} &= u_k - [G'_u(\theta_{k+1}, u_k)G'_u(\theta_{k+1}, u_k) + B'^*(\theta_{k+1}, u_k)B'(\theta_{k+1}, u_k)]^{-1} \\
&\quad \{G'^*(\theta_{k+1}, u_k)G(\theta_{k+1}, u_k) + B'^*(\theta_{k+1}, u_k)(B(\theta_{k+1}, u_k) - d)\}. \quad (20)
\end{aligned}$$

A simpler version of this algorithm was introduced and analyzed in [26]. In [26], the data-fitting operator,  $B$ , does not depend on the system parameter,  $\theta$ , and is a function of the state variable only, i.e.,  $B = B(u)$ . The IRGN scheme in Eq. (19) originates from variational regularization in the form

$$\min_{\theta \in \mathbb{R}^m} \left\{ \frac{1}{2} \|G(\theta, u_k)\|^2 + \frac{1}{2} \|B(\theta, u_k) - d\|^2 + \frac{\tau_k}{2} \|\theta - \bar{\theta}\|^2 \right\}. \quad (21)$$

The method in Eq. (20), on the other hand, is the classical Gauss–Newton algorithm applied to the nonlinear minimization problem

$$\min_{u \in \mathbb{R}^{4l}} \left\{ \frac{1}{2} \|G(\theta_{k+1}, u)\|^2 + \frac{1}{2} \|B(\theta_{k+1}, u) - d\|^2 \right\}. \quad (22)$$

The Gauss–Newton procedure in Eq. (20) does not need to be regularized, since solving the ODE system of equations in Eqs. (1)–(6), with respect to  $S$ ,  $V$ ,  $I_s$ ,  $I_v$ ,  $R$ , and  $D$ , is a forward problem, which is not generally ill-posed. Thus, its discrete approximation is also stable (as our numerical experiments below confirm).

The algorithm in Eqs. (19) and (20) was coded in MATLAB, using the optimization and parallel toolboxes. The code, along with figures, simulated data, and parameter estimates, can be found in our GitHub repository: <https://github.com/donajialej/WIMB2022team5.git>.

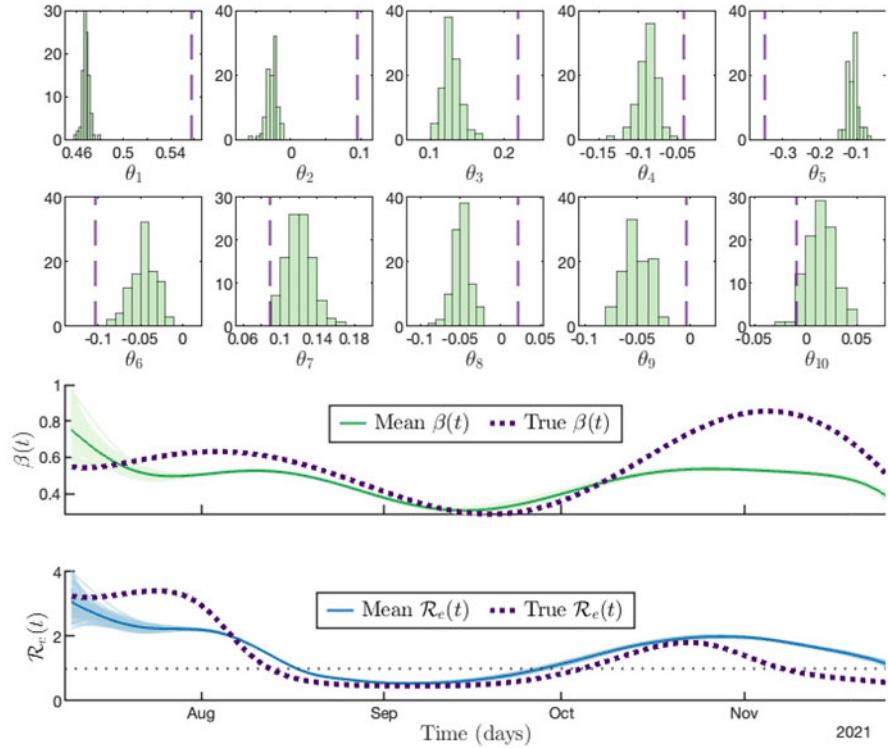
For all numerical simulations (with synthetic and real data), the unobserved state variables,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , are normalized; that is, in place of  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , we reconstruct the expansion coefficients for  $S/N$ ,  $V/N$ ,  $I_s/N$ , and  $I_v/N$ , where  $N$  is the total population of the region.

To select the number of basis functions for  $\beta(t)$  and for the unobserved state variables ( $m$  and  $n$ , respectively), we start with  $m = n = 5$  and keep increasing them until the reconstructed functions,  $\beta(t)$ ,  $S(t)$ ,  $V(t)$ ,  $I_s(t)$ , and  $I_v(t)$ , no longer visibly change.

An important part of parameter estimation is the choice of  $\lambda$  in Eqs. (10)–(11), which ensures that the two data sets—reported daily new cases and deaths—are well-balanced. In all our experiments, the value of  $\lambda$  is equal to 1000. For  $\lambda = 1$ , the misfit in daily new deaths is perceived as part of noise in incidence data, and the process is less sensitive to daily new deaths as compared to new incidence cases.

## 4 Numerical Experiments with Synthetic Data

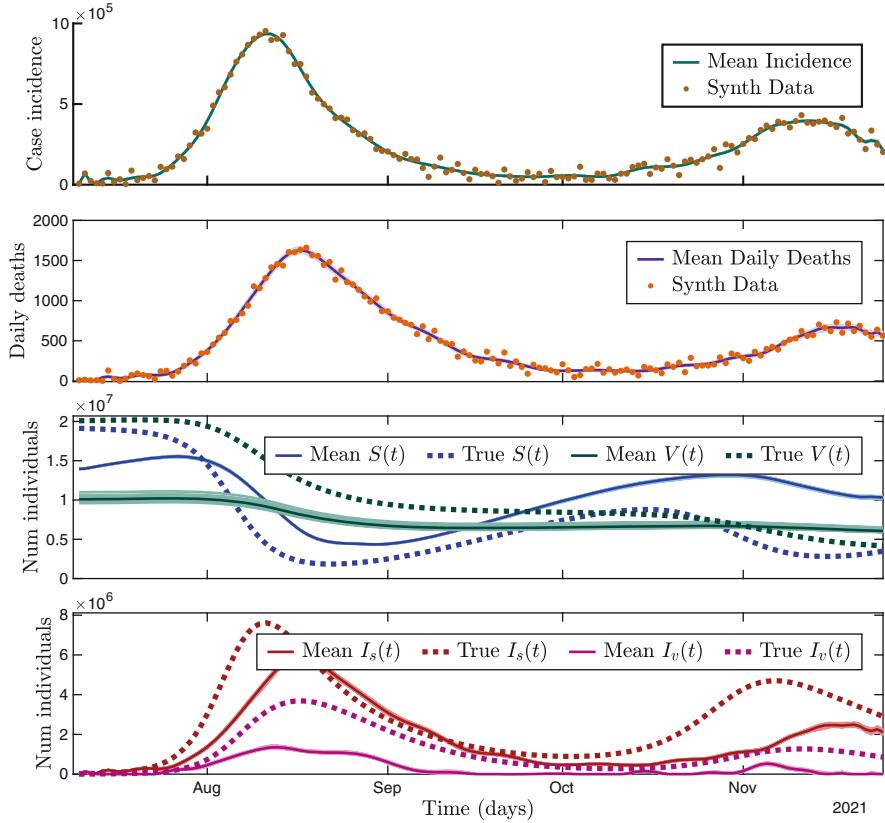
In this section, we test our proposed predictor–corrector algorithm (Eqs. (19)–(20)) using two synthetic data sets for incidence cases and deaths. The first synthetic data set was generated using the transmission rate  $\beta(t)$  shown in Fig. 2, which represents a case when initial success in disease prevention is followed by some setbacks causing the transmission rate to fluctuate. Specifically, this transmission rate was



**Fig. 2** Reconstruction of disease transmission  $\beta(t)$  (along with coefficients) and the effective reproduction number  $\mathcal{R}_e(t)$  for Scenario 1 (non-effective mitigation) from synthetic noisy data on new daily cases and deaths in Fig. 3. Simulations are carried out with 10 basis functions for the transmission rate  $\beta(t)$  and 40 basis functions for each unobserved state variable,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , i.e., 160 basis functions for all state variables combined. The regularization sequence is  $\tau_k = 10^{10}/(k+1)^{15}$ , and the iterations are stopped when  $k = 43$ . This stopping time is determined by the goodness of fit to both data sets

chosen to model a “non-effective mitigation” scenario where  $\mathcal{R}_e(t)$  remains above 1 for multiple time periods showing that the disease persists and spreads quickly. This is illustrated in the graph of  $\mathcal{R}_e(t)$  in Fig. 2. The second synthetic data set was generated using the transmission rate shown in Fig. 4 and represents an “effective mitigation” scenario where the disease transmission rate is reduced during the study period and where  $\mathcal{R}_e(t)$  stays below 1 more consistently.

In what follows, we evaluate the performance of the proposed method in reconstructing the unknown time-dependent transmission rate,  $\beta(t)$ , given synthetic daily rates of incidence cases and new deaths over a certain period of time. Two model transmission rates, described above, were selected (see Figs. 2 and 4). Each model transmission rate was used to solve the forward problem, i.e., the system of ODEs (Eqs. (1)–(6)), and to generate clean data on incidence cases,  $C(t)$ , and



**Fig. 3** Synthetic study of Scenario 1: non-effective mitigation. Top to bottom: synthetic (Synth) data (dots) and model fit (solid line) for daily new cases and daily new deaths; true synthetic values (dash line) and model reconstructions (solid line) for  $S(t)$  (blue),  $V(t)$  (green),  $I_s(t)$  (red), and  $I_v(t)$  (pink). There are 100 bootstrap model reconstructions, and the mean of them is a darker line of the color corresponding to each compartment

daily new deaths,  $\mathcal{T}(t)$ , on a given time interval  $[t_1, t_n]$  according to expressions, Eqs. (8) and (9), respectively. Then, random Gaussian noise (with 0 mean and a rather aggressive standard deviation) was added to epidemic data in order to mimic noise-contaminated data in a real-life setting, as shown in the top panels of Figs. 3 and 5. Since real incidence cases and deaths are known to be positive, uniform noise was added if the incidence became negative at any point.

Given “real” data for incidence cases and daily new deaths, we employed the regularized algorithm (Eqs. (19) and (20)) to simultaneously reconstruct the unknown transmission rate,  $\beta(t)$ , and the state variables,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , with discrete approximation given by Eqs. (14) and (15). In order to quantify uncertainty in the extracted transmission rate, we refit the model (using parallel programming via the *parfor* function in MATLAB) to  $M = 100$  additional data sets for incidence

cases and daily deaths assuming Poisson error structure. The resulting  $M$  best-fit parameter sets are used to build the histogram for each Legendre coefficient,  $\theta_j$ ,  $j = 1, 2, \dots, m$ , representing the frequency distribution of the reconstructed values.

To ensure an unbiased choice of the initial guess for  $\beta(t)$ , we take  $[\beta_0, 0, \dots, 0]^T$  to serve as initial approximation for  $[\theta_1, \theta_2, \dots, \theta_m]^T$  at every bootstrap iteration, where  $0.1 < \beta_0 < 1$ . To find initial approximations for  $u$ , we solve the system of ODEs (Eqs. (1)–(6)) with  $\beta(t) = \beta_0$  one time before the start of the iterative process and then evaluate Legendre expansion coefficients for the computed  $S, V, I_s$ , and  $I_v$  to form the initial vector  $u := [u_1, \dots, u_l, u_{l+1}, \dots, u_{2l}, u_{2l+1}, \dots, u_{3l}, u_{3l+1}, \dots, u_{4l}]^T$ .

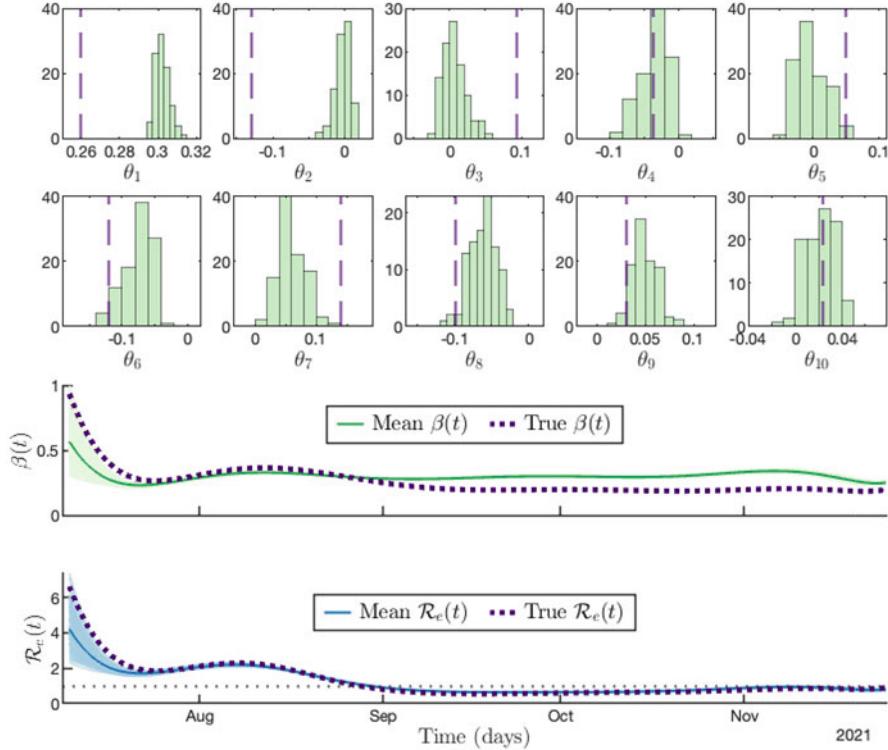
For the non-effective mitigation scenario (Scenario 1) with transmission rate  $\beta(t)$  shown in Fig. 2, the fitting procedure is initiated with  $\beta_0 = 0.5$  and is carried out using  $m = 10$  basis functions for the transmission rate,  $\beta(t)$ , and  $n = 40$  basis functions for each unobserved state variable,  $S, V, I_s$ , and  $I_v$ , giving a total of 160 basis functions for all state variables combined.

With no regularization, the iterative process to estimate the transmission rate in Scenario 1 (Fig. 2) turns out to be divergent. However, the process can be stabilized with a broad range of initial values,  $\tau_0$ , as long as they are consistent with the rate of decay of the regularization sequence,  $\tau_k$ . In our experiment, we selected  $\tau_0 = 10^{10}$  and the regularization sequence,  $\tau_k = 10^{10}/(k + 1)^{15}$ , the fastest rate of decrease that gives rise to a convergent iterative process. Iterations of Eqs. (19) and (20) are stopped when  $k = 43$ . This stopping time is determined by the goodness of fit to both data sets  $C$  and  $\mathcal{T}$ .

For the effective mitigation case (Scenario 2), where the transmission rate  $\beta(t)$  is presented in Fig. 4, the parameter estimation process is initiated with  $\beta_0 = 0.3$ . As before, the reconstruction is done with  $m = 10$ ,  $n = 40$ , and  $\tau_0 = 10^{10}$ , and the regularization sequence is driven to zero at the rate  $10^{10}/(k + 1)^{15}$ . But in this scenario, the iterative process is terminated when  $k = 19$ .

Figures 2 and 4 illustrate the connection between exact and reconstructed effective reproduction numbers,  $\mathcal{R}_e(t)$ , for the two scenarios with different model transmission rates. As stated in Sect. 2,  $\mathcal{R}_e(t) > 1$  describes time periods for which the disease persists and spreads quickly, and  $\mathcal{R}_e(t) < 1$  describes time periods for which the disease is contained (i.e., the disease is spreading slowly, eventually dying out). In the non-effective mitigation scenario described in Fig. 2, we see two approximately month-long windows for which the disease persists, highlighting that after the first push to decrease transmission ( $\mathcal{R}_e(t)$  falls to less than 1 in mid-August), mitigation strategies are not successful at keeping the transmission rate low enough, and a second wave begins in early October. For the effective mitigation scenario, described in Fig. 4, we see that although the effective reproduction rate  $\mathcal{R}_e(t)$  is greater than 1 for an extended initial period of time, once it drops below 1 (close to September) it stays below 1.

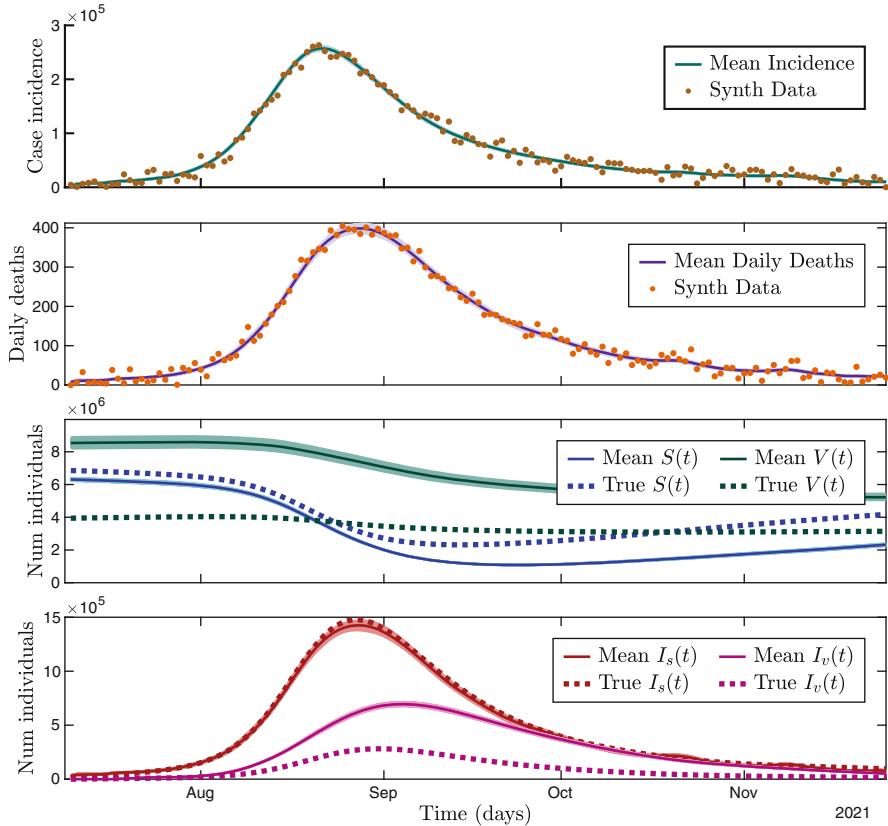
The top panels of Figs. 3 and 5 show how the bundles of incidence curves for daily new cases and deaths corresponding to the reconstructed transmission rates,  $\beta(t)$ , are compared to the noisy synthetic data used for data fitting.



**Fig. 4** Reconstruction of disease transmission  $\beta(t)$  (along with coefficients) and the effective reproduction number  $\mathcal{R}_e(t)$  in Scenario 2 (effective mitigation) from synthetic data on new daily cases and deaths in Fig. 5. Simulations are carried out with 10 basis functions for the transmission rate  $\beta(t)$  and 40 basis functions for each unobserved state variable,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , i.e., 160 basis functions for all state variables combined. The regularization sequence is  $\tau_k = 10^{10}/(k+1)^{15}$ , and the iterations are stopped when  $k = 19$ . This stopping time is determined by the goodness of fit to both data sets

Reconstructed  $S(t)$ ,  $V(t)$ ,  $I_s(t)$ , and  $I_v(t)$  from these two scenarios can be viewed in the lower panels of Figs. 3 and 5, respectively. While there are inevitable errors due to noise contamination in both data sets and due to accuracy loss stemming from regularization, Figs. 2, 3, 4, and 5 illustrate numerical experiments for synthetic data where the uncertainty is very low and the reconstruction of all unknown parameters is very stable. Yet, as evident from Figs. 3 and 5, it is harder to reconstruct the dynamics of the vaccinated population compared to the susceptible one since vaccinated individuals are less likely to contribute to new incidence cases (and especially deaths).

When comparing the time series for the reconstructed state variables between our two scenarios in the lower panels of Figs. 3 and 5, the progression of the disease follows the trend of the disease transmission rates. In particular, two infection peaks



**Fig. 5** Synthetic study of Scenario 2: effective mitigation. Top to bottom: synthetic (Synth) data (dots) and model fit (solid line) for daily new cases and daily new deaths; true synthetic values (dash line) and model reconstructions (solid line) for  $S(t)$  (blue),  $V(t)$  (green),  $I_s(t)$  (red), and  $I_v(t)$  (pink). There are 100 bootstrap model reconstructions, and the mean of them is a darker line of the color corresponding to each compartment

are in the lower panel of Fig. 3, which follow the peaks in the transmission rate and effective reproduction number curves in Fig. 2. A similar trend for a single infected peak is in the lower panel of Fig. 5, which follows the peaks in the transmission rate and effective reproduction number curves in Fig. 4. We also note that in the non-effective mitigation scenario (Fig. 3) the initial population is assumed to be  $N = 39,237,836$  and for the effective mitigation scenario (Fig. 5)  $N = 10,799,566$ .

Our simulated data and the inversion results for both experiments with synthetic data largely depend on the values of pre-estimated parameters,  $p$ ,  $\alpha$ ,  $\gamma_{s,r}$ ,  $\gamma_{v,r}$ ,  $\gamma_{s,d}$ ,  $\gamma_{v,d}$ ,  $\delta_v$ , and  $\delta_r$ , and the initial values for  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ . In both scenarios, we simulated for 140 days with the parameters as those from the real epidemic listed in Table 2. For initial values of  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , see the lower panels of Figs. 3 and 5.

## 5 Simulations with Real Data for COVID-19 Pandemic

In this section, we apply our SVIRD model (Eqs. (1)–(6)) and regularized computational algorithm (Eqs. (19) and (20)) to real data on incidence cases and new daily deaths for the second wave of COVID-19 in the United States in 2021, when the Delta variant was one of the more widely spread strains [36]. Most states experienced this second wave during an approximately 4-month period between July 9 and November 25, 2021, while vaccines were distributed to the US general population starting in early 2021. So we can study the progression of the pandemic under the effect of vaccination. For our experiments, we choose data sets for two states, Georgia and California, as both have different population sizes (Georgia is much smaller with approximately 11 million people versus the nearly 40 million living in California), had different proportions of vaccinated individuals between July 9 and November 25, 2021, and had different COVID-19 protocols. In particular, California had more vaccinated people at the onset and at the end of this time window [36], and California had stricter masking protocols; masks were required indoors in most places during this time period, whereas they were only recommended in the state of Georgia. The model variables and initial conditions corresponding to the population sizes in Georgia and California at the onset of the second wave are given in Table 1. Initial conditions were found using Census and CDC data [36–39]. Here,  $I(0) = I_s(0) + I_v(0)$  is the number of cases within the most recent week of the onset of the second wave, as most people with COVID-19 are no longer contagious 5 days after they first have symptoms and have been fever-free for at least 3 days.

**Table 1** Initial conditions used in the SVIRD model for the Georgia and California data. Population size was based on the January 7, 2021 data from <https://www.census.gov/quickfacts/GA> and <https://www.census.gov/quickfacts/CA>

Variable	Meaning	
$S(t)$	Number of susceptible unvaccinated individuals	
$V(t)$	Number of susceptible vaccinated individuals	
$I_s(t)$	Number of infectious unvaccinated individuals	
$I_v(t)$	Number of infectious vaccinated individuals	
$R(t)$	Number of recovered individuals	
$D(t)$	Number of deceased individuals	
Initial condition	Georgia	California
$S(0)$	$10,799,566 - V(0) - I(0)$	$39,237,836 - V(0) - I(0)$
$V(0)$	3,942,002	20,086,693
$I_s(0)$	3,580	25,039
$I_v(0)$	$731 (= 3580 * 5116 / 25039)$	5,116
$R(0)$	0	0
$D(0)$	0	0

**Table 2** Parameter values recorded for California and Georgia during the *second wave* of the pandemic, July 9–November 25, 2021 (approximately 4 months). The bars “–” in the last column mean that these values were calculated using  $\gamma_{s,d}$ , as described in the text

Parameter	Meaning	Value	Source
$\beta(t)$	Transmission rate		
$p$	Vaccination rate	0.00086 day <sup>-1</sup>	[37–39]
$\alpha$	Vaccine dose efficacy	0.8	[40, 43, 44]
$\gamma_{s,r}$	Recovery rate of unvaccinated	0.0995 day <sup>-1</sup>	–
$\gamma_{v,r}$	Recovery rate for vaccinated	0.09996 day <sup>-1</sup>	–
$\gamma_{s,d}$	Case-fatality for unvaccinated	0.00027 day <sup>-1</sup>	[45]
$\gamma_{v,d}$	Case-fatality for vaccinated	0.000021 day <sup>-1</sup>	–
$\delta_v$	Loss of immunity for vaccinated	0 day <sup>-1</sup>	[46]
$\delta_s$	Loss of immunity for unvaccinated	0.011 day <sup>-1</sup>	

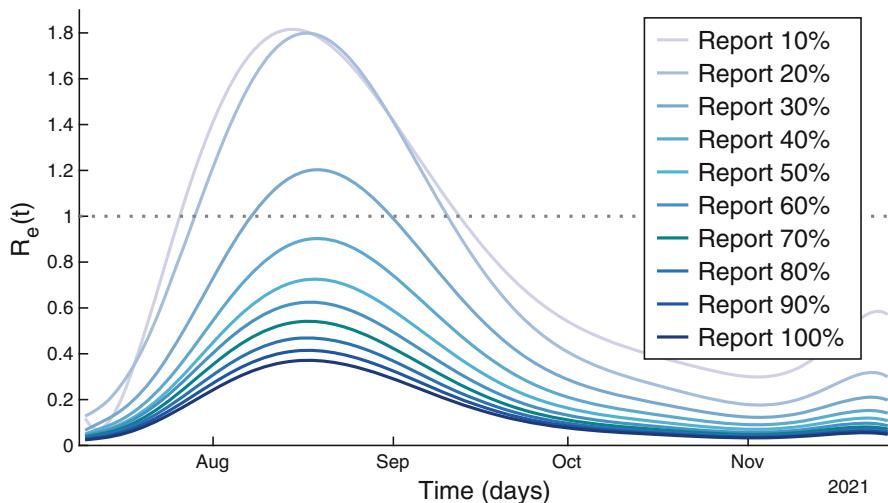
System parameter values used for California and Georgia during the *second wave* of the pandemic are presented in Table 2. The rationale for the selection of these values is as follows:

- Vaccination rate  $p$ : Based on the CDC data [39], during the selected time window, the proportion of fully vaccinated people changed from 37.5% to 49.8% in Georgia and from 51.1% to 63.1% in California, both of which resulted in about 12% increase in vaccination. Dividing this by our 140-day window gives the approximate daily vaccination rate  $p$  of 0.00086 day<sup>-1</sup>.
- Vaccine effectiveness  $\alpha$ : We choose  $\alpha = 0.8$  as the age-standardized crude vaccine effectiveness for infection was reported at 80% during July–November of 2021 [40].
- Death rate  $\gamma_{s,d}$ : We calculate  $\gamma_{s,d} = 0.005/18.5 = 0.00027$  days<sup>-1</sup> as the infectious fatality ratio IFR was reported as 0.5% from [41], and the median time from illness onset to death is 18.5 days (reported number for vaccinated vs unvaccinated [42]).
- Death rate  $\gamma_{v,d}$ : We take  $\gamma_{v,d} = (0.005/12.7)/18.5 = 0.000021$  days<sup>-1</sup> because during October–November, unvaccinated persons had 12.7 times the risks for COVID-19–associated death compared with those that were vaccinated without booster doses [33].
- Recovery rate  $\gamma_{s,r}$ : Assuming that individuals infected with COVID-19 either recover or die and using a recovery rate of 10 days, we conclude that the recovery rate for unvaccinated individuals is  $\gamma_{s,r} = (1 - 0.005)/10 = 0.0995$  days<sup>-1</sup>.
- Recovery rate  $\gamma_{v,r}$ : With a similar rationale as above, we estimate the recovery rate for vaccinated individuals as  $\gamma_{v,r} = (1 - 0.005/12.7)/10 = 0.09996$  days<sup>-1</sup>.
- Loss of immunity rate for recovered individuals  $\delta_r$ : We set  $\delta_s = 1/90 = 0.011$  days<sup>-1</sup>.
- Loss of immunity rate for vaccinated individuals  $\delta_v$ : We use  $\delta_v = 0$  as the Moderna and Pfizer-BioNTech vaccines offer immunity against COVID-19 for at least 6 months, and most people in the USA were fully vaccinated by the

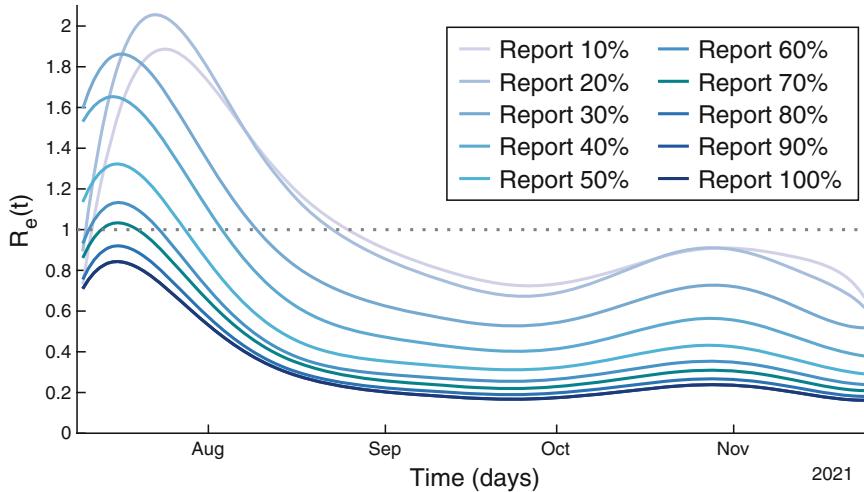
end of April 2021 or later. Therefore, they still had immunity against COVID-19 during most of the study period.

In the case of real data, apart from the measurement errors, which were incorporated in our earlier experiments, we also encounter modeling errors, which make the process considerably more unstable. Thus, apart from the penalty term,  $\frac{\tau_k}{2} \|\theta - \bar{\theta}\|^2$ , the iterative scheme also needs to be regularized by discretization. For this reason, fewer basis functions are used for the state variables. Specifically, we take 6 basis functions for each unobserved state variable,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , for the Georgia data, and 12 basis functions for each unobserved state variable for the California data. To further stabilize the process, we also introduce a smaller step size,  $\zeta = 0.1$ , as we update  $S(t)$ ,  $V(t)$ ,  $I_s(t)$ , and  $I_v(t)$ . This calls for more iterations needed to achieve the desirable data fit. The iterative process is terminated when  $k = 130$  for the Georgia data with regularization sequence  $\tau_k = 1/(k+1)^{10}$  and  $k = 58$  for the California data with  $\tau_k = 10^3/(k+1)^7$ . Overall, the time until convergence remains the same as for the case of synthetic data since the increase in the number of iterations is balanced by the reduction in the number of basis functions.

Another important aspect is the reporting rate of new cases. While it is natural to assume that the reporting rate for deaths due to COVID-19 is high, the reporting rate for daily new COVID-19 cases is unlikely to be anywhere close to 100% considering the large number of mild and asymptomatic cases (“silent spreaders” [47]). Figures 6



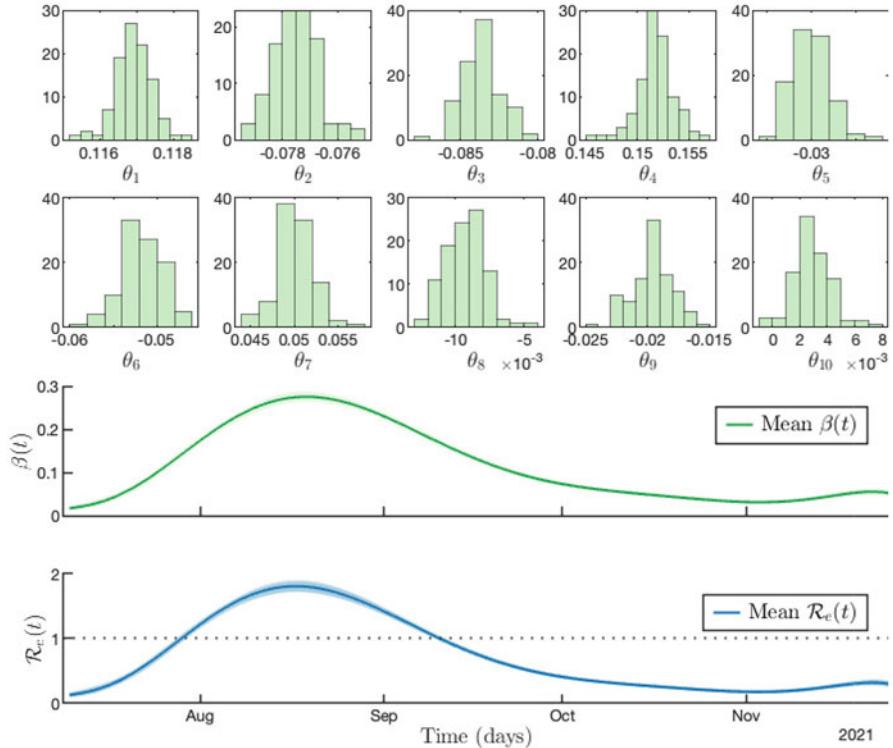
**Fig. 6** Reconstructed effective reproduction numbers,  $\mathcal{R}_e(t)$ , for various assumed reporting rates in the state of Georgia. Simulations are carried out with 10 basis functions for the transmission rate,  $\beta(t)$ , and 6 basis functions for each unobserved state variable,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , i.e., 24 basis functions for all state variables combined. The regularization sequence is  $\tau_k = 1/(k+1)^{10}$ , and the iterations are stopped when  $k = 130$ . This stopping time is determined by the goodness of fit to the Georgia data set



**Fig. 7** Reconstructed effective reproduction numbers,  $\mathcal{R}_e(t)$ , for various assumed reporting rates in the state of California. Simulations are carried out with 10 basis functions for the transmission rate,  $\beta(t)$ , and 12 basis functions for each unobserved state variable,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , i.e., 48 basis functions for all state variables combined. The regularization sequence is  $\tau_k = 10^3/(k+1)^7$ , and the iterations are stopped when  $k = 58$ . This stopping time is determined by the goodness of fit to the California data set

and 7 compare reconstructed time-dependent effective reproduction numbers,  $\mathcal{R}_e(t)$ , for various assumed reporting rates of daily new cases in Georgia and California, respectively (for both states, we fixed the reporting rate for daily new deaths due to COVID-19 at 90%). We know that at the onset of the Delta variant wave of the COVID-19 pandemic, the reproduction number must have been above 1 for some time. Thus, Fig. 6 suggests that the reporting rate of new COVID-19 incidence cases in the state of Georgia is 10–30%. For California, we see that the reporting rate is 10–60% as illustrated in Fig. 7. This is consistent with the estimation of COVID-19 incidence reporting rate carried out in [48]. In [48], the reporting rate was cast as one of the unknown parameters in the model and had to be reconstructed by the optimization algorithm. For the initial pre-vaccination stage of COVID-19 pandemic in the state of Georgia, the reporting rate for new incidence cases was estimated to be 0.23 (95% confidence interval (CI): [0.22, 0.24]). For the reasons listed previously and as suggested by our numerical study, in simulations presented in Figs. 8, 9, 10, and 11, we assume a 90% reporting rate for new daily deaths due to COVID-19 and a 20% reporting rate for new incidence cases in the states of Georgia and California.

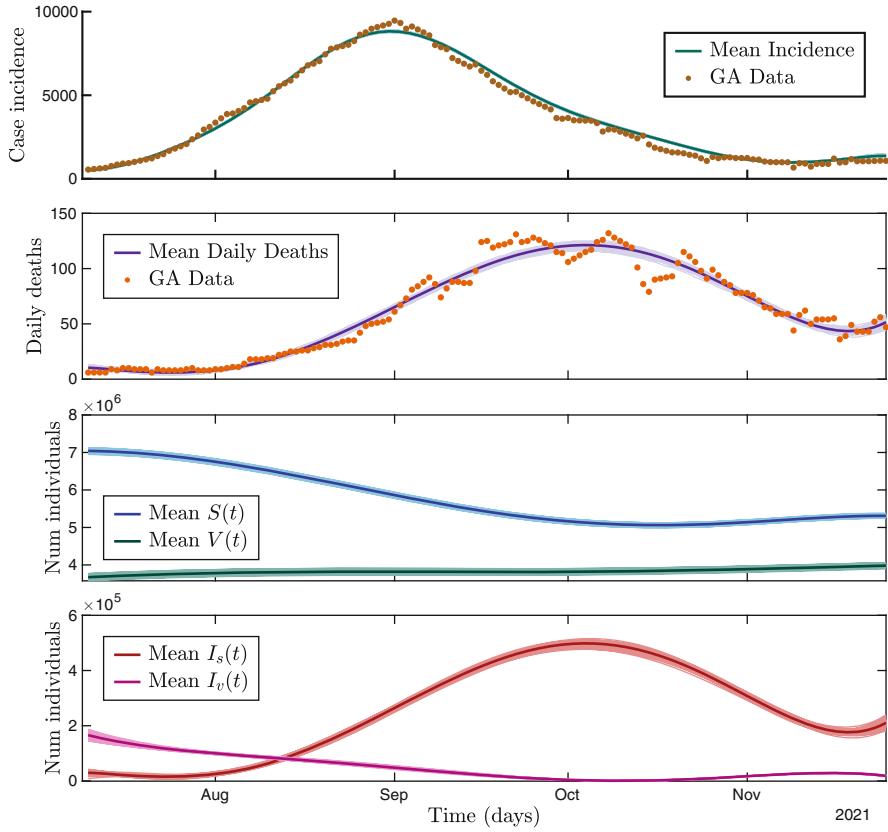
In Figs. 8 and 10, we show the transmission rate,  $\beta(t)$ , and the effective reproduction number,  $\mathcal{R}_e(t)$ , reconstructed from daily data on new cases and deaths for the states of Georgia and California, respectively, for the period from July 9 to November 25, 2021. The top panels of Figs. 9 and 11 show how incidence curves for daily new cases and deaths in the states of Georgia and California are compared to real data used for parameter estimation in the optimization process (Eqs. (19)



**Fig. 8** Reconstruction of disease transmission  $\beta(t)$  (along with coefficients) and the effective reproduction number  $\mathcal{R}_e(t)$  for the state of Georgia

and (20)). Reconstructed  $S(t)$ ,  $V(t)$ ,  $I_s(t)$ , and  $I_v(t)$  for the states of Georgia and California can be viewed in the lower panel of the same figures. One may notice that the California incidence data (top panel of Fig. 11) are more “spread out” than the Georgia incidence data (top panel of Fig. 9). This is because, for the Georgia data, a rolling 7-day average was recorded each week since in Georgia new cases were often not reported on the weekends when the Delta variant was dominant. So, the approximation of unobserved state variables for the state of California is more uncertain as compared to Georgia and to the sets of synthetic data.

The parameter estimation process is initiated with  $\beta_0 = 0.5$  for both Georgia and California. The reconstruction is done with  $m = 10$  in both cases (the number of basis functions for the transmission rate). For Georgia, the number of basis functions for each unobserved state variable is  $n = 6$  (i.e., 24 basis functions for all state variables,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , combined). The iterative process started with  $\tau_0 = 1$ . The regularization sequence is driven to zero at the rate  $1/(k+1)^{10}$ . Like in the case of Georgia, for the California data set the number of basis functions for  $S$ ,  $V$ ,  $I_s$ , and  $I_v$  is significantly reduced (from  $n = 40$  to  $n = 12$ ), as compared to

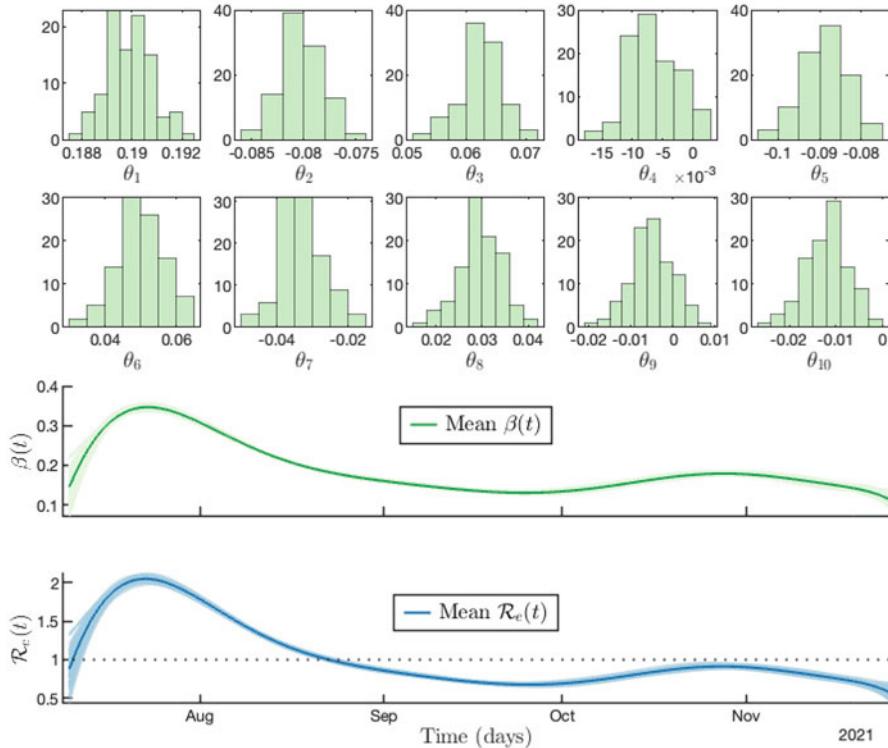


**Fig. 9** State of Georgia (GA) case study. Top to bottom: state data (dots) and model fit (solid line) for daily new cases and daily new deaths; 100 bootstrap model reconstructions for  $S(t)$  (blue),  $V(t)$  (green),  $I_s(t)$  (red), and  $I_v(t)$  (pink). The mean of the bootstraps is a darker line of the color corresponding to each compartment

reconstructions with synthetic data in order to further stabilize predictor–corrector algorithm (Eqs. (19) and (20)) in the presence of modeling error.

By comparing Figs. 8 and 10, one can see that the start of the Delta variant wave in the state of California was more rapid as compared to Georgia, but it took longer for Georgia to get the virus under control (as compared to California). In California, the effective reproduction number,  $\mathcal{R}_e(t)$ , dropped under 1 around mid-August, while in Georgia  $\mathcal{R}_e(t)$  remained greater than 1 until early September 2021. However, in California, the effective reproduction number almost bounced back to 1 in late October before going down again toward the end of the study period. In Georgia, on the other hand,  $\mathcal{R}_e(t)$  remained very low after the end of September.

In the top panels of Figs. 9 and 11, we note the peak of around 9,000 new incidence cases in the state of Georgia in early September and the peak in mid-August of approximately 13,000 new incidence cases in the state of California.

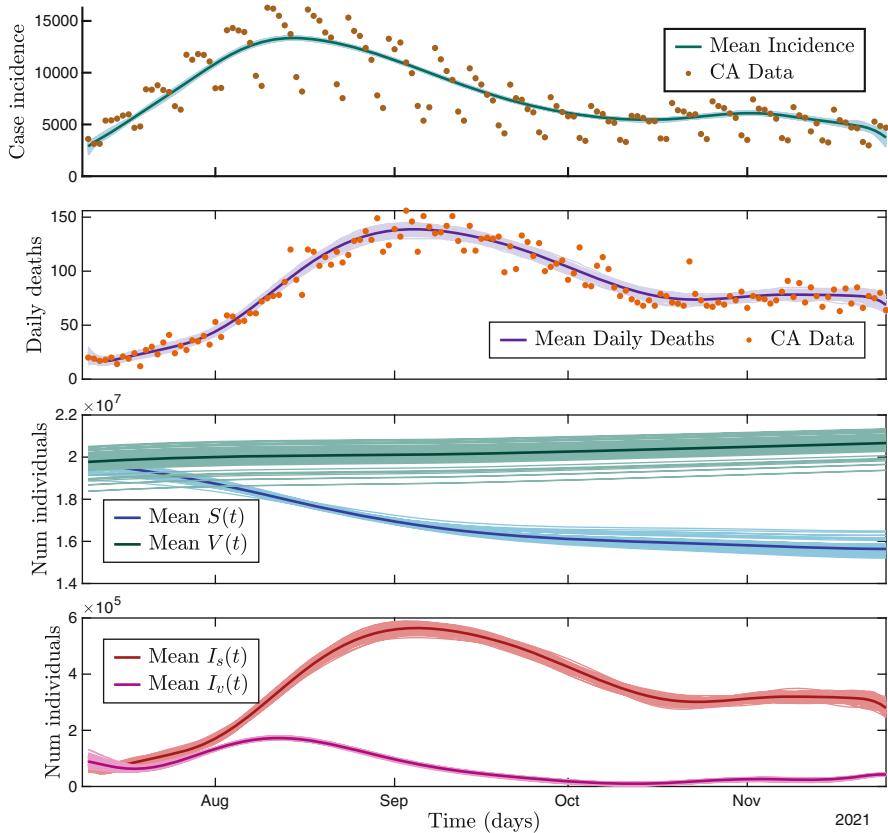


**Fig. 10** Reconstruction of disease transmission  $\beta(t)$  (along with coefficients) and the effective reproduction number  $\mathcal{R}_e(t)$  for the state of California

In both states the daily reported new deaths are under 150 people. The peaks in deaths follow the peaks of incidence cases, in early October in Georgia and in early September in California. Reconstructed curves,  $I_s(t)$  and  $I_v(t)$ , are consistent with the reported percentage of vaccinated individuals in the states of Georgia and California, respectively (Figs. 9 and 11).

## 6 Conclusion and Future Work

In this chapter, we propose a new dynamic model of COVID-19 transmission that takes into account the vaccination status of both susceptible and infected humans. It also includes a possible loss of immunity and reinfection within both vaccinated and unvaccinated populations. To estimate the unknown disease parameters, we develop a novel computational algorithm, which employs a parameter cascade approach. The proposed method is used to reconstruct time-dependent transition rates,  $\beta(t)$ ,



**Fig. 11** State of California (CA) case study. Top to bottom: state data (dots) and model fit (solid line) for daily new cases and daily new deaths; 100 bootstrap model reconstructions for  $S(t)$  (blue),  $V(t)$  (green),  $I_s(t)$  (red), and  $I_v(t)$  (pink). The mean of the bootstraps is a darker line of the color corresponding to each compartment

and effective reproduction numbers,  $\mathcal{R}_e(t)$ , from synthetic and real data for the COVID-19 pandemic. Apart from COVID-19, the proposed compartmental model and iteratively regularized optimization method can be applied to the study of other infectious diseases.

In the course of our numerical study, the new optimization technique has emerged as a reliable alternative to more traditional trust-region and gradient-descent algorithms that are commonly used in parameter estimation. The efficiency of these algorithms is limited when a complex biological model (which may be a system of nonlinear ordinary or partial differential equations) constraining the underlying minimization problem does not have a closed-form solution and has to be solved numerically at every step of the iterative process. Our new method, on the other hand, does not require either exact or approximate solution to the constraining system.

In reconstructing time-dependent transmission rates,  $\beta(t)$ , in order to reduce the computational load and to improve the estimate efficiency, we pre-specified the values of other system parameters by conducting a thorough review of the literature. To assess the sensitivity of reconstructed transmission rates to slight variations in pre-estimated parameters, one can build a Bayesian model to assign priors to pre-specified parameters, and the posterior distributions of transmission rates will incorporate the uncertainty in these parameters. This is an important topic for future work. Note that for a simpler SIRD model corresponding to a pre-vaccination stage of the COVID-19 pandemic, the sensitivity analysis has been conducted in [48]. In [48], for every bootstrap iteration, the recovery rate,  $\gamma$ , and the fatality rate,  $\nu$ , have been sampled from normal distributions,  $N(0.20, 0.02)$  and  $N(0.005, 0.001)$ , respectively. The normal distribution,  $N(0.20, 0.02)$ , for the recovery rate,  $\gamma$ , reflected an average infectious period between 3 and 20 days, while the normal distribution  $N(0.005, 0.001)$  for the fatality rate,  $\nu$ , accounted for the variation of this parameter within different risk groups. The reconstructed values of  $\beta(t)$  with normally distributed  $\gamma$  and  $\nu$  were almost identical to those reconstructed with constant (mean) values of these pre-estimated parameters showing a very low sensitivity of  $\beta(t)$  to inevitable variations in COVID-19 infectious periods and fatality rates.

With a considerable portion of mild and asymptomatic cases, the number of reported daily new cases is much lower than the actual value. In this chapter, we change the reporting rates of new incidence cases and investigate how different reporting rates affect the reconstruction of effective reproduction numbers,  $\mathcal{R}_e(t)$ , in our numerical simulations. Thus, another important direction of future research will be to modify our reconstruction process to include the estimation of the unknown percentages of new incidence cases along with the unknown time-dependent transmission rate,  $\beta(t)$ , and other system parameters. The problem of the reporting rate can also be addressed by extending the model to include the compartment of asymptomatic spreaders.

We also plan to add line search routines and incorporate nonnegativity constraints for unobserved state variables,  $S$ ,  $V$ ,  $I_s$ , and  $I_v$ , in iteratively regularized predictor–corrector algorithm (Eqs. (19) and (20)). This will allow further accuracy improvements and stability of the proposed optimization method.

Last but not least, the methodology must be extended to provide near real-time forecasting of future incidence cases and deaths (among vaccinated and unvaccinated individuals) from early data for an unfolding outbreak. This research is crucial for control and prevention, in particular, for the assessment of various vaccination strategies.

**Acknowledgments** The work described herein was initiated during the Collaborative Workshop for Women in Mathematical Biology funded and hosted by UnitedHealth Group Optum of Minnetonka, MN and supported by University of Minnesota’s Institute for Mathematics and its Applications in June 2022. Additionally, the authors and editors thank the anonymous peer reviewers for their feedback, which strengthened this work.

AS was supported by the National Science Foundation (NSF) under Grant No. DMS-2011622.

## References

1. A. Smirnova, G. Chowell, *Infect. Dis. Model.* **2**(2), 268 (2017)
2. W. Kermack, A. McKendrick, *Proc. R. Soc. London Ser. A* **115**(772), 700 (1927)
3. N.T. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications* (Charles Griffin & Company, High Wycombe, 1975)
4. R. Schlickeiser, M. Kroger, *Physics* **3**(2), 386 (2021)
5. Y. Mohamadou, A. Halidou, P. Kapen, *Appl. Intell.* **50**(11), 3913 (2020)
6. A. Mahajan, N. Sivadas, R. Solanki, *Chaos Solitons Fractals* **140**, 110156 (2020)
7. R. Singh, R. Adhikari, Preprint. arXiv:2003.12055 (2020)
8. T. Sardar, S. Nadim, S. Rana, J. Chattopadhyay, *Chaos Solitons Fractals* **139**, 110078 (2020)
9. M. Dalton, P. Dougall, F. Amoah-Darko, W. Annan, E. Asante-Asamani, S. Bailey, J. Greene, D. White, *PLoS ONE* **17**(11), e0274407 (2022). <https://doi.org/10.1371/journal.pone.0274407>
10. W. Wong, F. Juwono, T. Chua, Preprint. arXiv:2101.07494 (2021)
11. M. Angeli, G. Neofotistos, M. Mattheakis, E. Kaxiras, *Chaos Solitons Fractals* **154**, 111621 (2022)
12. C. Kirkeby, T. Halasa, M. Gussmann, *Sci. Rep.* **7**(9496) (2017)
13. D. Bates, D. Watts, *Nonlinear Regression Analysis and its Applications* (Wiley, Hoboken, 1988)
14. G. Seber, C. Wild, *Nonlinear Regression* (Wiley, Hoboken, New Jersey, 2003)
15. J. Cao, J. Huang, H. Wu, *J. Comput. Gr. Stat.* **21**(1), 42 (2012)
16. J. Ramsay, G. Hooker, *Dynamic Data Analysis* (Springer, New York, 2017)
17. J.M. Varah, *SIAM J. Sci. Comput.* **3**(1), 28 (1982)
18. H. Liang, H. Wu, *J. Am. Stat. Assoc.* **103**(484), 1570 (2008)
19. J. Ramsay, G. Hooker, D. Campbell, J. Cao, *J. R. Stat. Soc. Ser. B* **69**(5), 741 (2007)
20. H. Attouch, J. Bolte, P. Redont, A. Soubeyran, *Math. Oper. Res.* **35**(2), 438 (2010)
21. J. Bolte, S. Sabach, M. Teboulle, *Math. Program.* **146**, 459 (2014)
22. M. Hong, Z. Luo, M. Razaviyayn, *SIAM J. Optim.* **26**(1), 2016)
23. A. Patrascu, I. Necoara, *J. Global Optim.* **61**(1), 19 (2015)
24. D. Driggs, J. Tang, J. Liang, M. Davies, C. Schönlieb, *SIAM J. Imaging Sci.* **14**(4) (2021)
25. D. Davis, M. Udell, B. Edmunds, *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems* (2016), pp. 226–234
26. A. Smirnova, A. Bakushinsky, *Inverse. Prob.* **36**(12), 125015 (2020)
27. M. Eisenberg, S. Robertson, J. Tien, *J. Theo. Biol.* **324**, 84 (2013)
28. K. Roosa, G. Chowell, *Theo. Biol. Med. Model.* **16**(1), 1 (2019)
29. N. Tuncer, H. Gulbudak, V. Cannataro, M. Martcheva, *Bull. Math. Biol.* **9**(78), 1796 (2016)
30. G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A.D. Filippo, A.D. Matteo, M. Colaneri, *Nat. Med.* **26**, 855 (2020)
31. K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J.M. Hyman, P. Yan, G. Chowell, *J. Clin. Med.* **9**, 596 (2020)
32. R.N. Thompson, Epidemiological models are important tools for guiding COVID-19 interventions, *BMC Med.* **18**, 152 (2020). <https://doi.org/10.1186/s12916-020-01628-4>
33. A.G. Johnson, A.B. Amin, A.R. Ali, et al., COVID-19 incidence and death rates among unvaccinated and fully vaccinated adults with and without booster doses during periods of Delta and Omicron variant emergence - 25 U.S. Jurisdictions, April 4–December 25, 2021. *Morb Mortal Wkly Rep (MMWR)* **71**, 132–138 (2022). <http://doi.org/10.15585/mmwr.mm7104e2>
34. O. Diekmann, J. Heesterbeek, J. Metz, *J. Math. Biol.* **28**(4), 365 (1990)
35. P. van den Driessche, J. Watmough, *Math. Biosci.* **180**(1–2), 29 (2002)
36. Centers for Disease Control and Prevention. Trends in Number of COVID-19 Cases and Deaths in the US Reported to CDC, by State/Territory. [https://covid.cdc.gov/covid-data-tracker/#trends\\_dailycases\\_select\\_06](https://covid.cdc.gov/covid-data-tracker/#trends_dailycases_select_06) (2022). Accessed 30 Sep 2022

37. America Counts Staff. California Remained Most Populous State but Growth Slowed Last Decade. <https://www.census.gov/library/stories/state-by-state/california-population-change-between-census-decade.html>. Accessed 30 Sep 2022
38. America Counts Staff. Georgia Among Top Five Population Gainers Last Decade. <https://www.census.gov/library/stories/state-by-state/georgia-population-change-between-census-decade.html> (2021). Accessed 30 Sep 2022
39. Centers for Disease Control and Prevention. Trends in Number of COVID-19 Vaccinations in the US. <https://covid.cdc.gov/covid-data-tracker/#vaccination-trends> (2022). Accessed 30 Sep 2022
40. Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report. <https://www.cdc.gov/mmwr/volumes/71/wr/mm7104e2.htm> (2022). Accessed 30 Sep 2022
41. M. O'Driscoll, G.R. dos Santos, L. Wang, D.A.T. Cummings, A.S. Azman, J. Paireau, A. Fontanet, S. Cauchemez, H. Salje, *Nature* **590**, 140–145 (21)
42. Drugs.com. How Do COVID-19 Symptoms Progress and What Causes Death? <https://www.drugs.com/medical-answers/covid-19-symptoms-progress-death-3536264/> (2022). Accessed 30 Sep 2022
43. N. Lewis, L. Chambers, H. Chu, T. Fortnam, R.D. Vito, L. Gargano, P. Chan, J. McDonald, J. Hogan, *JAMA Netw. Open*, 1–11 (2022). <https://doi.org/10.1001/jamanetworkopen.2022.23917>
44. J. Lopez-Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrerera, et al., *N. Engl. J. Med.* **385**(7), 585 (2021)
45. UCHealth. The Delta Variant of COVID-19. <https://www.uchealth.org/services/infectious-diseases/coronavirus-covid-19/the-delta-variant-of-covid-19/> (2022). Accessed 30 Sep 2022
46. N. Doria-Rose, M.S. Suthar, M. Makowski, S. O'Connell, A.B. McDermott, B. Flach, J.E. Ledgerwood, J.R. Mascola, B.S. Graham, B.C. Lin, *N. Engl. J. Med.* **384**(23), 2259 (2021)
47. K. Hwang, C. Edholm, O. Saucedo, L. Allen, N. Shakiba, *Bull. Math. Biol.* **84**(9), 91 (2022)
48. A. Smirnova, B. Pidgeon, R. Luo, J. Inverse. Ill-Posed. Prob. **30**(6), 823 (2022). <https://doi.org/10.1515/jiip-2021-0037>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

