Research Article

Alexandra Smirnova*, Brian Pidgeon and Ruiyan Luo

On stable parameter estimation and short-term forecasting with quantified uncertainty with application to COVID-19 transmission

https://doi.org/10.1515/jiip-2021-0037
Received June 22, 2021; revised February 2, 2022; accepted March 7, 2022

Abstract: A novel optimization algorithm for stable parameter estimation and forecasting from limited incidence data for an emerging outbreak is proposed. The algorithm combines a compartmental model of disease progression with iteratively regularized predictor-corrector numerical scheme aimed at the reconstruction of case reporting ratio, transmission rate, and effective reproduction number. The algorithm is illustrated with real data on COVID-19 pandemic in the states of Georgia and New York, USA. The techniques of functional data analysis are applied for uncertainty quantification in extracted parameters and in future projections of new cases.

Keywords: Regularization, forecasting, parameter estimation

MSC 2010: 47J06, 65J20, 65L08

1 Introduction

Emerging and re-emerging infectious diseases continue to generate significant morbidity and mortality around the world [13]. Globally, as of December 6, 2021, the ongoing COVID-19 outbreak has brought about 265,194,191 confirmed cases, including 5,254,116 deaths [35, 36]. Although many people infected with COVID-19 have mild or no symptoms, the virus can also cause severe, even fatal illness.

Since COVID-19 is a new virus, researchers around the world learn more about it every day. Despite much success in the application of mathematical and statistical tools to our understanding of COVID-19 progression, the key challenge remains stable estimation of important disease parameters, such as reproductive capacity of the virus and its underlying transmission rate [9, 20, 25]. Estimating these parameters early on allows for the real-time analysis of the effectiveness of control and prevention and enables accurate forecasting of future incidence cases [12]. Whereas other system parameters, i.e., incubation and recovery rates, are less dependent on public health policies, the effective reproduction number and the transmission rate of the disease (denoted by $\mathcal{R}(t)$ and $\beta(t)$, respectively) 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, particularly in the face of noise-contaminated data and substantial uncertainty in approximate solutions [28].

***Corresponding author: Alexandra Smirnova,** Department of Mathematics & Statistics, Georgia State University, Atlanta, USA, e-mail: asmirnova@gsu.edu
Brian Pidgeon, Department of Mathematics & Statistics, Georgia State University, Atlanta, USA, e-mail: bpidgeon1@student.gsu.edu
Ruiyan Luo, Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, USA, e-mail: rluo@gsu.edu

In this study, we present a numerical algorithm for stable estimation of the effective reproduction number $\mathcal{R}(t)$ and disease transmission rate $\beta(t)$ from COVID-19 data (early pre-vaccination stage) on confirmed new cases and deaths, using the following version of the SIRD model introduced in [2]:

$$\frac{dS}{dt} = -\beta(t)\frac{S(t)}{\mathcal{N} - D(t)}I(t), \tag{1.1}$$

$$\frac{dI}{dt} = \beta(t)\frac{S(t)}{\mathcal{N} - D(t)}I(t) - \gamma I(t), \tag{1.2}$$

$$\frac{dR}{dt} = (1-\nu)\gamma I(t), \tag{1.3}$$

$$\frac{dD}{dt} = \nu\gamma I(t). \tag{1.4}$$

It is important to mention that our *parameter estimation/forecasting* algorithm is not limited to any particular compartmental model and/or to any particular virus. In our study, we employ incidence data for the early stage of COVID-19 pandemic to illustrate the efficiency of the proposed method. The regularized numerical method that we develop can be used for a broad class of models and for various emerging outbreaks. At the early stage, little is known about the virus and our understanding of the virus is changed every day. With limited information about an emerging disease, it is advisable to use a relatively simple model that does not incorporate too many assumptions that may or may not prove to be true.

Equations (1.1)–(1.4) follow the progression of individuals in a population of size N between four different states: S, susceptible to the COVID-19 virus, I, infected with COVID-19 (both symptomatic and asymptomatic), R, recovered and no longer contagious, and D, deceased. The parameter γ, called *recovery rate*, governs the evolution of infected people from state I to states R or D. In line with [2], we assume that the mean value of γ is $1/5$, which corresponds to the infectious period of 5 days. To account for the possibility of the infectious period to range between 3 to 20 days [30], we conduct a sensitivity analysis with randomly selected values of γ from a normal distribution, $N(0.20, 0.02)$.

The second parameter ν is the *fatality rate* of the virus. Estimating the fatality rate is extremely difficult since COVID-19 cases are believed to be grossly underreported (in part, due to a large number of asymptomatic cases, especially among children and young adults, and at a later stage among vaccinated individuals). While early measurements from limited data for an emerging COVID-19 outbreak suggested ν to be as high as 1.2 %, the more recent estimates based on antibody testing point towards a much lower value of 0.2 % (though it does increase markedly with age and risk factors). In our numerical simulations, we assume the mean value of ν to be 0.5 % [2]. For the sensitivity analysis, we randomly sample the value of this parameter from a normal distribution, $N(0.005, 0.001)$, to account for variation within different risk groups.

The *transmission rate* $\beta(t)$ is defined as probability of infection given a contact between an infectious and susceptible individual multiplied by the average rate of contacts between these groups. It is the defining rate in evolution of any disease and one of the two components in the *effective reproduction number* $\mathcal{R}(t)$, the rate at which susceptible agents get infected divided by the recovery rate of infected individuals at time t (see [2]), i.e.,

$$\mathcal{R}(t) = \frac{\beta(t)}{\gamma}\frac{S(t)}{\mathcal{N} - D(t)}. \tag{1.5}$$

The effective reproduction number of a disease varies in time, and it is directly affected by social response and public health guidelines, by which it eventually falls under 1 for a sustained period of time needed to stop the chain of transmission.

The paper is organized as follows. In Section 2, the constrained nonlinear minimization problem aimed at the estimation of COVID-19 transmission rate $\beta(t)$ and effective reproduction number $\mathcal{R}(t)$ is formulated. In Section 3, a modified version of the iteratively regularized predictor-corrector algorithm [22] for solving the constrained least squares problem (CLSP) is introduced, and some unique features of Jacobian and Hessian approximations of the nonlinear operator Φ in the CLSP are studied in Section 4. Numerical simulations on stable estimation of disease parameters from full data sets and forecasting of new incidence cases and daily deaths from partial data are presented in Sections 6 and 7, respectively, with uncertainty quantification

routine outlined in Section 5. All experiments are carried out with real data on the COVID-19 pandemic in the states of Georgia and New York, USA. Conclusions and future plans are discussed in Section 8.

2 The constrained nonlinear minimization problem

Let $d^{(1)}$ and $d^{(2)}$ be incidence data on new COVID-19 confirmed cases and deaths, respectively. Naturally, we assume that both data sets are noise contaminated and

$$\max\{\|d^{(1)} - d_\delta^{(1)}\|,\ \|d^{(2)} - d_\delta^{(2)}\|\} \le \delta,$$

where $\|\cdot\|$ is the Euclidian norm in $\mathbb{R}^n$ and n is the number of data points in each set. Note that, according to (1.2), the daily number of new COVID-19 cases is (see [2])

$$\beta(t)\frac{S(t)}{\mathcal{N} - D(t)}I(t) = \frac{dI}{dt} + \gamma I(t). \tag{2.1}$$

Therefore, multiplying the right-hand side of (2.1) by ψ, the unknown reporting rate, we get the number of cases that are actually reported. On the other hand, by (1.4), the daily number of new deaths is $\nu\gamma I(t)$, and we suppose here that for COVID-19 related deaths the reporting rate is 100 %. Assume that, in a particular region, the first COVID-19 case is reported on day t_1, while t_n is the last day of the study period. Let the data for new cases and deaths, $d_\delta^{(1)}$ and $d_\delta^{(2)}$, be reported on days $t_1, t_2, \dots, t_n$. Then our goal is to recover the unknown time-dependent transmission rate while solving the following constrained minimization problem:

$$\begin{aligned}\min_{\psi,I} f(\psi, I), \quad f(\psi, I) &:= \frac{\lambda_1}{2}\left\|\psi\Big(\frac{dI}{dt} + \gamma I\Big) - d^{(1)}\right\|^2 + \frac{\lambda_2}{2}\|\nu\gamma I - d^{(2)}\|^2 \\ &= \frac{\lambda_1}{2}\sum_{i=1}^{n}\Big(\psi\Big(\frac{dI}{dt} + \gamma I\Big)(t_i) - d_i^{(1)}\Big)^2 + \frac{\lambda_2}{2}\sum_{i=1}^{n}(\nu\gamma I(t_i) - d_i^{(2)})^2\end{aligned} \tag{2.2}$$

subject to (1.1)–(1.4). In (2.2), the role of λ_1 and λ_2 is to balance the two terms in the cost functional in order to ensure that the two data sets $d^{(1)}$ and $d^{(2)}$ play an equal part in minimization. From (2.2), one obtains

$$f(\psi, I) = \frac{\lambda_1}{2}\psi^2\left\|\Big(\frac{dI}{dt} + \gamma I\Big)\right\|^2 - \psi\lambda_1\Big(\frac{dI}{dt} + \gamma I, d^{(1)}\Big) + \frac{\lambda_1}{2}\|d^{(1)}\|^2 + \frac{\lambda_2}{2}\|\nu\gamma I - d^{(2)}\|^2.$$

In the above, $(\cdot, \cdot)$ is the scalar product in the Euclidian space $\mathbb{R}^n$. By the first-order necessary condition, one concludes

$$\frac{1}{\lambda_1}\frac{\partial f}{\partial \psi} = \psi\left\|\Big(\frac{dI}{dt} + \gamma I\Big)\right\|^2 - \Big(\frac{dI}{dt} + \gamma I, d^{(1)}\Big) = 0,$$

which implies

$$\psi = \frac{\big(\frac{dI}{dt} + \gamma I, d^{(1)}\big)}{\big\|\big(\frac{dI}{dt} + \gamma I\big)\big\|^2}. \tag{2.3}$$

Substituting (2.3) into (2.2), one arrives at the following least squares problem:

$$\min_I\left\{\frac{\lambda_1}{2}\left\|\frac{\big(\frac{dI}{dt} + \gamma I, d^{(1)}\big)}{\big\|\big(\frac{dI}{dt} + \gamma I\big)\big\|^2}\Big(\frac{dI}{dt} + \gamma I\Big) - d^{(1)}\right\|^2 + \frac{\lambda_2}{2}\|\nu\gamma I - d^{(2)}\|^2\right\} \tag{2.4}$$

subject to (1.1)–(1.4). Suppose that $t = a$ is one day before the first case is reported, that is,

$$a < t_1 < t_2 < \cdots < t_n < b \quad \text{and} \quad I'(a) = I(a) = 0.$$

Denote $W(t) := I'(t)$, and discretize $W(t)$ by using the Fourier approximation

$$W(t) = A_0 + \sum_{j=1}^{N}\Big\{A_j\cos\Big(2\pi j\frac{t-a}{b-a}\Big) + B_j\sin\Big(2\pi j\frac{t-a}{b-a}\Big)\Big\}.$$

To ensure that $W(a) = W(b) = 0$, we set $A_0 = -\sum_{j=1}^{N} A_j$. This yields

$$W(t) = \sum_{j=1}^{N} \left\{ A_j \left[\cos\left(2\pi j \frac{t-a}{b-a}\right) - 1 \right] + B_j \sin\left(2\pi j \frac{t-a}{b-a}\right) \right\}. \tag{2.5}$$

For $I(t)$, (2.5) implies

$$I(t) = \int_a^t W(s)\, ds = \sum_{j=1}^{N} \left\{ A_j \left[\frac{b-a}{2\pi j} \sin\left(2\pi j \frac{t-a}{b-a}\right) - (t-a) \right] - B_j \frac{b-a}{2\pi j} \left[\cos\left(2\pi j \frac{t-a}{b-a}\right) - 1 \right] \right\}, \tag{2.6}$$

and from (2.6), it follows that

$$U(t) := \int_a^t I(s)\, ds = \sum_{j=1}^{N} \left\{ A_j \left[-\frac{(b-a)^2}{(2\pi j)^2} \left\{ \cos\left(2\pi j \frac{t-a}{b-a}\right) - 1 \right\} - \frac{(t-a)^2}{2} \right] - B_j \frac{b-a}{2\pi j} \left[\frac{b-a}{2\pi j} \sin\left(2\pi j \frac{t-a}{b-a}\right) - (t-a) \right] \right\}. \tag{2.7}$$

Furthermore, to recover the shape of $\beta(t)$ in (1.1)–(1.2), we project the transmission rate onto a finite subset spanned by shifted Legendre polynomials of degree $0, 1, \ldots, m-1$, which are orthogonal on the interval $[a, b]$ with respect to the L_2 inner product, defined recursively as follows:

$$x = \frac{2t-a-b}{b-a}, \quad P_0(x) = 1, \quad P_1(x) = x, \quad t \in [a, b],$$
$$(j+1)P_{j+1}(x) = (2j+1)xP_j(x) - jP_{j-1}(x), \quad j = 1, 2, \ldots, m-2.$$

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

$$\beta(t) = \sum_{j=0}^{m-1} \theta_{j+1} P_j(t). \tag{2.8}$$

Define the vector of unknown state variables as

$$u := [A_1, \ldots, A_N, B_1, \ldots, B_N]^T,$$

and introduce the following operators:

$$B_i(u) := W[u](t_i) + \gamma I[u](t_i), \quad \psi(u) := \frac{(B(u), d^{(1)})}{\|(B(u))\|^2}, \quad \Phi_i(u) := \psi(u) B_i(u),$$
$$\Omega_i(u) := \nu\gamma I[u](t_i), \quad i = 1, 2, \ldots, n. \tag{2.9}$$

According to (1.1)–(1.4),

$$\beta(t) \frac{S(t)}{\mathcal{N} - D(t)} I(t) = \beta(t) \frac{N - I(t) - \gamma U(t)}{N - \nu\gamma U(t)} I(t). \tag{2.10}$$

Thus, if one sets

$$G_i(\theta, u) := B_i(u) - \beta[\theta](t_i) \frac{N - I[u](t_i) - \gamma U[u](t_i)}{N - \nu\gamma U[u](t_i)} I[u](t_i) \quad g := 0, \quad i = 1, 2, \ldots, n, \tag{2.11}$$

then one arrives at the following constrained minimization problem:

$$\min_{u \in \mathbb{R}^{2N}} \left\{ \frac{\lambda_1}{2} \|\Phi(u) - d^{(1)}\|^2 + \frac{\lambda_2}{2} \|\Omega(u) - d^{(2)}\|^2 \right\} \tag{2.12}$$

$$\text{subject to} \quad G(\theta, u) = g, \quad G\colon \mathbb{R}^m \times \mathbb{R}^{2N} \to \mathbb{R}^n. \tag{2.13}$$

3 Iteratively regularized predictor-corrector algorithm

Fitting model predictions to aggregated time series of case incidence and deaths yields an ill-posed problem due to excessive noise propagation coupled with substantial underreporting of epidemic data (due to a large number of asymptomatic and mild cases). In order to solve this ill-posed problem in a stable manner, regularized optimization algorithms [3, 8, 10, 11, 24, 26, 27] are commonly used to minimize the cost functional. Oftentimes, practical implementation of these algorithms consists in reducing (2.4), constrained by (1.1)–(1.4), to the least squares problem over the parameter space only,

$$\min_{\beta}\Big\{\frac{\lambda_1}{2}\|J_1(\beta)-d^{(1)}\|^2+\frac{\lambda_2}{2}\|J_2(\beta)-d^{(2)}\|^2\Big\}. \tag{3.1}$$

To arrive at (3.1), one solves the system of differential equations (1.1)–(1.4) for the parameter-to-state map $[S, I, R, D] = [S[\beta](t), I[\beta](t), R[\beta](t), D[\beta](t)]$ and then substitutes $I = I[\beta](t)$ into (2.4). Once (3.1) has been derived and discretized, the unknown parameter $\beta(t)$ is computed by a regularized Gauss–Newton, Levenberg–Marquardt, or a gradient-type algorithm. Thus, in (3.1), J_1 and J_2 are compositions of $I = I[\beta](t)$, satisfying (1.1)–(1.4), and the observation operators

$$\frac{(\frac{dI}{dt}+\gamma I, d^{(1)})}{\|(\frac{dI}{dt}+\gamma I)\|^2}\Big(\frac{dI}{dt}+\gamma I\Big) \quad \text{and} \quad \nu\gamma I,$$

respectively. The ODE system (1.1)–(1.4) is nonlinear. Therefore, at every step of the iterative process, the state variable $I_k = I[\beta_k](t)$ needs to be calculated numerically for each current value of β_k. This can noticeably increase the computational complexity of parameter estimation, while at the same time making it even more sensitive to the presence of noise in the input data.

In this paper, for the recovery of $\beta(t)$, we implement a regularized predictor-corrector algorithm, similar to the one developed in [22], which avoids solving (1.1)–(1.4) at every step of the iterative process, and by doing so, it incorporates an extra layer of stability in the inversion procedure. In this algorithm, one discretizes both the unknown parameter and the state variable as proposed in the previous section. In order to solve the resulting finite-dimensional problem (2.12)–(2.13), one updates θ while freezing u, and then u is modified while θ is kept unchanged. More specifically, given $\binom{\theta_k}{u_k}$, one transitions from θ_k to θ_{k+1} by applying one step of the preconditioned iteratively regularized Gauss–Newton (PIRGN) scheme [3, 11, 21, 24]

$$\theta_{k+1}=\theta_k-[G'^{*}_{\theta}(\theta_k,u_k)G'_{\theta}(\theta_k,u_k)+\alpha_k T^*T]^{-1}\{G'^{*}_{\theta}(\theta_k,u_k)(G(\theta_k,u_k)-g)+\alpha_k T^*T(\theta_k-\bar{\theta})\}. \tag{3.2}$$

Then, given $\binom{\theta_{k+1}}{u_k}$, one uses the classical Gauss–Newton procedure [15, 16] to update u_k,

$$\begin{aligned}u_{k+1}=u_k-&[G'^{*}_{u}(\theta_{k+1},u_k)G'_{u}(\theta_{k+1},u_k)+\lambda_1\Phi'^{*}_{2}(u_k)\Phi'_{2}(u_k)+\lambda_2\Omega'^{*}(u_k)\Omega'_{2}(u_k)]^{-1}\\&\times\{G'^{*}_{u}(\theta_{k+1},u_k)(G(\theta_{k+1},u_k)-g)+\lambda_1\Phi'^{*}_{2}(u_k)(\Phi(u_k)-d^{(1)}_{\delta})+\lambda_2\Omega'^{*}(u_k)(\Omega(u_k)-d^{(2)}_{\delta})\}.\end{aligned} \tag{3.3}$$

Note that PIRGN scheme (3.2) originates from variational regularization [26, 29] in the form

$$\min_{\theta\in\mathbb{R}^m}\Big\{\frac{1}{2}\|G(\theta,u_k)-g\|^2+\frac{\alpha_k}{2}\|T(\theta-\bar{\theta})\|^2\Big\},$$

where $\frac{\alpha_k}{2}\|T(\theta-\bar{\theta})\|^2$ is a stabilizing penalty term and $\{\alpha_k\}$ is a regularization sequence that tends to zero as k approaches infinity. In practice, we terminate iterations (3.2)–(3.3) at an appropriate stopping time $k = \mathcal{K}(\delta)$ to ensure convergence. For iteration (3.2) to be well-defined, we assume that T is a surjective linear operator between two Hilbert spaces $\mathbb{R}^m$ and $\mathcal{X}$ satisfying the following condition [24]: for any $h \in \mathbb{R}^m$,

$$(T^*Th,h)\geq\zeta\|h\|^2, \quad \zeta>0.$$

In the above, $\mathcal{X}$ is either finite or infinite-dimensional. As an example, T can be the operator mapping $[\theta_1,\theta_2,\dots,\theta_m]^T$ to $\beta[\theta](t)=\sum_{j=0}^{m-1}\theta_{j+1}P_j(t)$, where $\{P_j(t)\}$ is the set of Legendre polynomials of degree

$0, 1, \dots, m-1$. The operator T can also be used to scale selected components of the unknown vector θ if they are expected to be of different orders of magnitude [24].

Method (3.3), on the other hand, is the classical Gauss–Newton algorithm applied to the nonlinear minimization problem

$$\min_{u\in\mathbb{R}^{2N}}\Big\{\frac{1}{2}\|G(\theta_{k+1},u)-g\|^2+\frac{\lambda_1}{2}\|\Phi(u)-d^{(1)}\|^2+\frac{\lambda_2}{2}\|\Omega(u)-d^{(2)}\|^2\Big\},$$
$$\max\{\|d^{(1)}-d_\delta^{(1)}\|,\ \|d^{(2)}-d_\delta^{(2)}\|\}\le\delta, \tag{3.4}$$

with one exception: in (3.3), the operator $\Phi'(u_k)$ is replaced with $\Phi_2'(u_k)$. The definition of $\Phi_2'(u_k)$ and the rationale for the replacement will be given in Section 4 below. Note that Gauss–Newton procedure (3.3) does not need to be regularized since minimizing the functional $\|G(\theta_{k+1},u)-g\|^2$ with respect to $u\in\mathbb{R}^{2N}$ is not an ill-posed problem: it is a forward problem in ordinary differential equations. The reader may consult [22] for the detailed convergence analysis of a predictor-corrector algorithm, which is a slightly simplified version of iterative scheme (3.2)–(3.3) used here.

4 Jacobian and Hessian approximation for $\Phi(u)$

In order to enhance stability of our optimization algorithm, in this section, we take a close look at the gradient of the functional $h(u):=\frac{1}{2}\|\Phi(u)-d^{(1)}\|^2$. According to (2.9),

$$\nabla h(u)=\Phi'^*(u)(\Phi(u)-d^{(1)})=\Phi'^*(u)(\psi(u)B(u)-d^{(1)}). \tag{4.1}$$

The Jacobian $\Phi'(u)$ can be expressed as the sum of two matrices $\Phi_1'(u)+\Phi_2'(u)$ in the following manner:

$$\Phi'(\mathbf{q})=\Big[\frac{\partial\psi}{\partial u_1}B(u)\quad\cdots\quad\frac{\partial\psi}{\partial u_{2N}}B(u)\Big]+\psi(u)\Big[\frac{\partial B}{\partial u_1}\quad\cdots\quad\frac{\partial B}{\partial u_{2N}}\Big]:=\Phi_1'(u)+\Phi_2'(u). \tag{4.2}$$

Since $\psi(u):=\frac{(B(u),d^{(1)})}{\|(B(u))\|^2}$, one has

$$\frac{\partial\psi}{\partial u_j}=\frac{(\frac{\partial B}{\partial u_j},d^{(1)})\|B(u)\|^2-2(\frac{\partial B}{\partial u_j},B(u))(B(u),d^{(1)})}{\|B(u)\|^4}$$
$$=\frac{(\frac{\partial B}{\partial u_j},d^{(1)})-2\psi(u)(\frac{\partial B}{\partial u_j},B(u))}{\|B(u)\|^2},\quad j=1,2,\dots,2N.$$

It follows from (4.1) and (4.2) that

$$\nabla h(u)=\Phi'^*(u)(\Phi(u)-d^{(1)})=(\Phi_1'(u)+\Phi_2'(u))^*(\Phi(u)-d^{(1)}).$$

Identity (4.2) implies that, for every $j=1,2,\dots,2N$,

$$[\Phi_1'^*(u)(\Phi(u)-d^{(1)})]_j=\Big(\frac{\partial\psi}{\partial u_j}B(u),\psi(u)B(u)-d^{(1)}\Big)=\frac{\partial\psi}{\partial u_j}\{\psi(u)\|B(u)\|^2-(B(u),d^{(1)})\}$$
$$=\frac{\partial\psi}{\partial u_j}\Big\{\frac{(B(u),d^{(1)})}{\|(B(u))\|^2}\|B(u)\|^2-(B(u),d^{(1)})\Big\}=0.$$

Hence, the residual $\Phi(u)-d^{(1)}$ is in the kernel of matrix $\Phi_1'(u)$. This yields a simplified expression for the gradient

$$\nabla h(u)=\Phi_2'^*(u)(\Phi(u)-d^{(1)})$$

with a reduced number of operations and, therefore, a reduced noise propagation due to unnecessary rounding. In consequence, we arrive at the following Hessian for the functional $h(u)$ (see [23]):

$$H(u)=\Phi_2'^*(u)\Phi'(u)+(\Phi_2''(u)\cdot)^*(\Phi(u)-d^{(1)})=\Phi_2'^*(u)(\Phi_1'(u)+\Phi_2'(u))+(\Phi_2''(u)\cdot)^*(\Phi(u)-d^{(1)}).$$

Assuming a reasonably small residual and to ensure that the Hessian approximation is symmetric and non-negative definite, in our predictor-corrector algorithm (3.2)–(3.3), we approximate $H(u)$ by

$$H(u) \approx \Phi_2'^{*}(u)\Phi_2'(u).$$

In the next sections, we present the results of our numerical simulations (with quantified uncertainty) from full data sets for COVID-19 incidence cases and deaths in the states of Georgia and New York, USA.

5 Uncertainty quantification

The goal of our experiments is to estimate the disease transmission rate $\beta(t)$ by solving the discretized minimization problem (2.12)–(2.13) with predictor-corrector algorithm (3.2)–(3.3) over the interval $[t_1, t_n]$, and then to use the reconstructed function $\beta(t)$ to calculate $\mathcal{R}(t)$ by formulas (1.5), (2.10), (2.8), (2.6), and (2.7).

To quantify uncertainty in the extracted reproduction numbers, we refit the model to $M = 100$ additional data sets for incidence cases and deaths (assuming Poisson error structure) [4, 6, 7], while sampling γ and ν from normal distributions, $N(0.20, 0.02)$ and $N(0.005, 0.001)$, respectively, and using uniform distribution on $[0.1, 1]$ for $\beta_0(t)$. This results in M approximate reproduction curves $\mathcal{R}_i(t)$, $i = 1, 2, \ldots, M$. We treat each reproduction curve as its own observation, and we use the technique of functional data analysis [17, 18] to compute 95 % pointwise confidence intervals for the mean function $\mathcal{R}(t)$ at every moment of time within $[t_1, t_n]$, i.e., such confidence intervals that the true reproduction number $\mathcal{R}(t)$ falls within the interval 95 % of time for any t_j, $j = 1, \ldots, n$. Let y_{ij} denote the observation from the reproduction curve $\mathcal{R}_i(t)$ at the time $t = t_j$. Consider the model

$$y_{ij} = \mathcal{R}(t_j) + \epsilon_i(t_j), \quad i = 1, \ldots, M,\ j = 1, \ldots, n,$$

where $\mathcal{R}(t)$ is the mean function and the errors $\epsilon_i(t)$ are independent for different i and can have within curve correlation, that is, the correlation between $\epsilon_i(t)$ and $\epsilon(t')$ for any t and t' in the domain. Introduce the basis expansion of $\mathcal{R}(t)$ in the following form:

$$\mathcal{R}(t) = \sum_{k=1}^{K} c_k \phi_k(t) = c'\phi,$$

where $\phi = (\phi_1(t), \ldots, \phi_K(t))'$ denotes the vector of cubic spline basis functions and $K = n + 2$. Using the second-order derivatives to measure smoothness, we estimate the coefficient vector c by solving the penalized least square problem

$$\min_c \sum_{i=1}^{M}\sum_{j=1}^{n}\{y_{ij} - c'\phi(t_j)\}^2 + \kappa \int_{t_1}^{t_n}\{\mathfrak{D}^2 c'\phi(t)\}^2\, dt,$$

where $\sum_{i=1}^{M}\sum_{j=1}^{n}\{y_{ij} - c'\phi(t_j)\}^2$ is the sum of squared residuals, $\kappa \int_{t_1}^{t_n}\{\mathfrak{D}^2 c'\phi(t)\}^2\, dt$ is the smoothness penalty, and $\mathfrak{D}^2 c'\phi$ denotes the second derivative of $c'\phi$. The estimate $\hat{c}$ of the coefficient vector c is given by

$$\hat{c} = (\mathfrak{F}'\mathfrak{F} + \kappa P)^{-1}\mathfrak{F}'\bar{y},$$

and the mean function is estimated as

$$\hat{\mathcal{R}}(t) = \hat{c}'\phi(t),$$

where $\mathfrak{F}$ denotes the $n \times K$ matrix of the values of basis functions with the (i, j) element $\phi_i(t_j)$, P denotes the $K \times K$ matrix with the (i, j) element $\int \mathfrak{D}^2\phi_i(t)\mathfrak{D}^2\phi_j(t)dt$, and $\bar{y} = (\bar{y}(t_1), \ldots, \bar{y}(t_n))'$, $\bar{y}(t_j) = \sum_{i=1}^{M} y_{ij}/M$. The covariance matrix of $\hat{c}$ is

$$\mathrm{Var}(\hat{c}) = (\mathfrak{F}'\mathfrak{F} + \kappa P)^{-1}\mathfrak{F}\Sigma_e\mathfrak{F}(\mathfrak{F}'\mathfrak{F} + \kappa P)^{-1}/M$$

where Σ_e is the population variance-covariance matrix of the residual vector ϵ and is estimated by the covariance matrix E of residuals by

$$\hat{\Sigma}_e = (M - 1)^{-1}E'E$$

Then the variance of the reproduction number is

$$\operatorname{var}(\widehat{\mathcal{R}(t)}) = \phi(t)' \operatorname{Var}(\hat{c})\phi(t).$$

Based on the Central Limit Theorem, with $M = 100$, we assume that the estimated mean function at each time point approximately has normal distribution. Therefore, we calculate 95 % pointwise confidence intervals for the mean function by adding and subtracting 2 standard errors from the mean function. In other words, our 95 % pointwise confidence intervals for each fixed time point are calculated by

$$(\hat{\mathcal{R}}(t) - 2\sqrt{\operatorname{var}(\widehat{\mathcal{R}(t)})}, \hat{\mathcal{R}}(t) + 2\sqrt{\operatorname{var}(\widehat{\mathcal{R}(t)})}).$$

The algorithm is implemented in the R package fda, and the default value for κ, $\kappa = 3 \cdot 10^{-8}/(t_n - t_1)$, which depends on the range of argument for the functions, is used.

6 Estimation of $\beta(t)$ and $\mathcal{R}(t)$ from full data sets

For successful implementation of algorithm (3.2)–(3.3), one needs to find initial approximations for the Legendre coefficients θ and for the Fourier coefficients u, used to discretize the unknown transmission rate $\beta(t)$ and the unknown state variable $I'(t)$, respectively. To ensure an unbiased choice of the initial guess for $\beta(t)$, we randomly select a constant θ_1^0 from the uniform distribution on $[0.1, 1]$ and take $\theta^0 = [\theta_1^0, 0, 0, \dots, 0]^T$ to serve as initial approximation for the transmission rate expansion coefficients at every bootstrap iteration. Note that this choice of θ^0 yields $\beta_0(t) = \theta_1^0$.

To find an initial guess for u, we over-fit the death wave by comprising u^0 of coefficients A_j^0 and B_j^0, $j = 1, 2, \dots, N$, that satisfy

$$\mathcal{D}(t) = \sum_{j=1}^{N} \left\{ A_j^0 \left[\frac{b-a}{2\pi j} \sin\left(2\pi j \frac{t-a}{b-a}\right) - (t-a) \right] - B_j^0 \frac{b-a}{2\pi j} \left[\cos\left(2\pi j \frac{t-a}{b-a}\right) - 1 \right] \right\}.$$

Here $\mathcal{D}(t)$ is a cubic spline interpolation of the vector $d_\delta^{(2)}/(\gamma\nu)$ on the interval $[a, b]$. With this approach, both θ^0 and u^0 will change with each additional data set used for uncertainty quantification. If one denotes $v = 2\pi(t-a)/(b-a)$ and multiplies both sides of the above identity by $\cos(mv)$, $m = 1, 2, \dots, N$, then one obtains

$$\int_0^{2\pi} \mathcal{D}\left(a + \frac{b-a}{2\pi} v\right) \cos(mv)\, dv = \sum_{j=1}^{N} \left\{ A_j^0 \frac{b-a}{2\pi} \int_0^{2\pi} (\sin(jv)/j - v) \cos(mv)\, dv - B_j^0 \frac{b-a}{2\pi j} \int_0^{2\pi} (\cos(jv) - 1) \cos(mv)\, dv \right\}.$$

This yields the following equation for B_m^0:

$$\int_0^{2\pi} \mathcal{D}\left(a + \frac{b-a}{2\pi} v\right) \cos(mv)\, dv = -B_m^0 \frac{b-a}{2\pi m} \int_0^{2\pi} \cos^2(mv)\, dv.$$

Thus, for $m = 1, 2, \dots, N$, one gets

$$B_m^0 = -\frac{4\pi m}{(b-a)^2} V_m, \quad V_m := \int_a^b \mathcal{D}(t) \cos\left(\frac{2\pi m(t-a)}{b-a}\right) dt.$$

Furthermore, if one uses the same substitution, $v = 2\pi(t-a)/(b-a)$, and multiplies the expression for $\mathcal{D}(t)$ by $\sin(mv)$, $m = 1, 2, \dots, N$, then one concludes

$$\int_0^{2\pi} \mathcal{D}\left(a + \frac{b-a}{2\pi} v\right) \sin(mv)\, dv = \sum_{j=1}^{N} \left\{ A_j^0 \frac{b-a}{2\pi} \int_0^{2\pi} (\sin(jv)/j - v) \sin(mv)\, dv - B_j^0 \frac{b-a}{2\pi j} \int_0^{2\pi} (\cos(jv) - 1) \sin(mv)\, dv \right\}. \tag{6.1}$$

Identity (6.1) implies the following equation for A_j^0, $j = 1, 2, \ldots, N$:

$$\int_0^{2\pi} \mathcal{D}\Big(a + \frac{b-a}{2\pi} v\Big) \sin(mv)\, dv = \frac{b-a}{2\pi}\Big(\frac{A_m^0}{m} \int_0^{2\pi} \sin^2(mv)\, dv - \sum_{j=1}^N A_j^0 \int_0^{2\pi} v \sin(mv)\, dv \Big)$$
$$= \frac{b-a}{m}\Big[\frac{A_m^0}{2} + \sum_{j=1}^N A_j^0 \Big].$$

Introduce the notation $Y_m := \int_0^{2\pi} \mathcal{D}(a + \frac{b-a}{2\pi} v) \sin(mv)\, dv$. Then

$$\frac{mY_m}{b-a} = \frac{A_m^0}{2} + \sum_{j=1}^N A_j^0 \quad \text{and} \quad \frac{1}{b-a} \sum_{m=1}^N mY_m = \Big(\frac{1}{2} + N\Big) \sum_{j=1}^N A_j^0, \tag{6.2}$$

and therefore, $\sum_{j=1}^N A_j^0 = \frac{2}{(b-a)(1+2N)} \sum_{j=1}^N jY_j$. If one plugs the expression for $\sum_{j=1}^N A_j^0$ into the first identity (6.2), then one has

$$\frac{mY_m}{b-a} = \frac{A_m^0}{2} + \frac{2}{(b-a)(1+2N)} \sum_{j=1}^N jY_j. \tag{6.3}$$

From (6.3), one concludes

$$A_m^0 = \frac{2}{b-a}\Big[mY_m - \frac{2}{1+2N} \sum_{j=1}^N jY_j \Big].$$

Going back to the variable t, one arrives at

$$A_m^0 = \frac{4\pi}{(b-a)^2}\Big[mZ_m - \frac{2}{1+2N} \sum_{j=1}^N jZ_j \Big], \quad Z_m := \int_a^b \mathcal{D}(t) \sin\Big(\frac{2\pi m(t-a)}{b-a} \Big)\, dt.$$

Next, given (2.9) and (2.11), we calculate Fréchet derivatives of $\Phi(u)$, $\Omega(u)$, and $G(\theta, u)$ with respect to u and the derivative of $G(\theta, u)$ with respect to θ. As justified in Section 4, we then replace $\Phi'(u)$ in (3.3) with its truncated version $\Phi_2'(u)$, given unique properties of the nonlinear operator $\Phi(u)$.

For our first numerical experiment, we take data on incidence cases and deaths for COVID-19 virus in the state of Georgia, USA, over the period of 256 days, from February 29, 2020, when the first two cases were reported, to November 10, 2020 (see Figures 1 and 2 for the data used and for the major public health

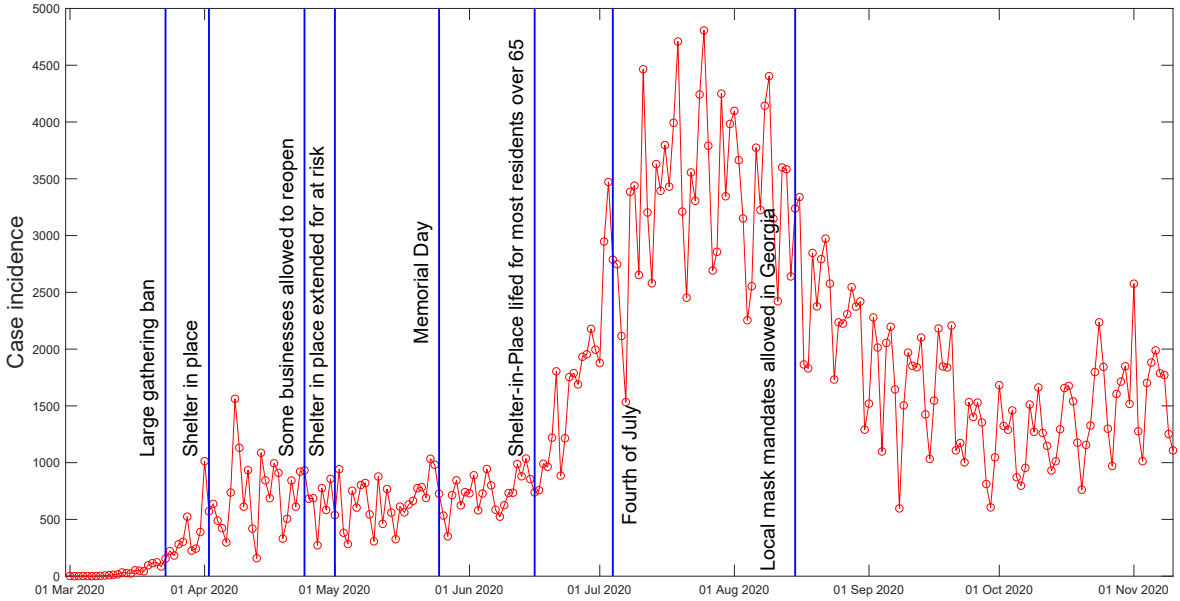

Figure 1: The state of Georgia, USA, COVID-19 incidence wave, 2020

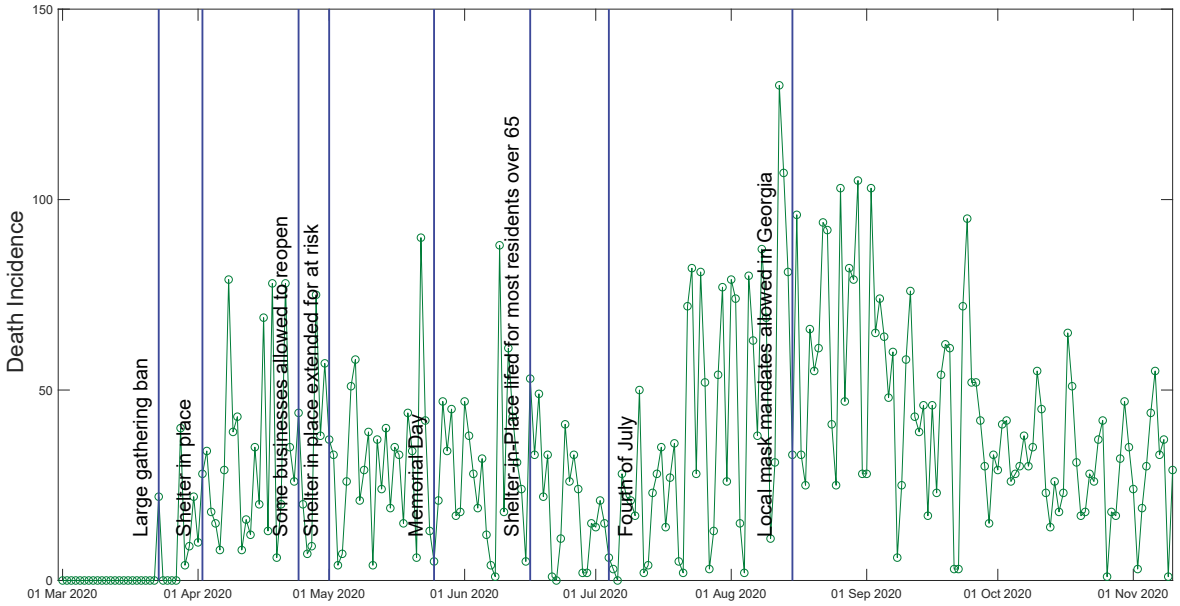

Figure 2: The state of Georgia, USA, COVID-19 death wave, 2020

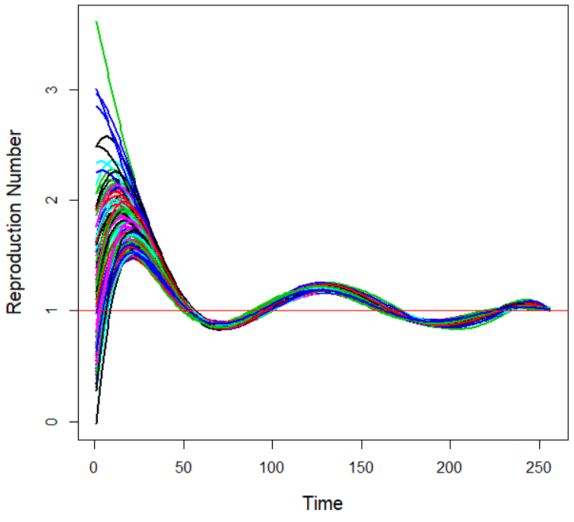

Figure 3: Georgia, USA: effective reproduction number as functional data

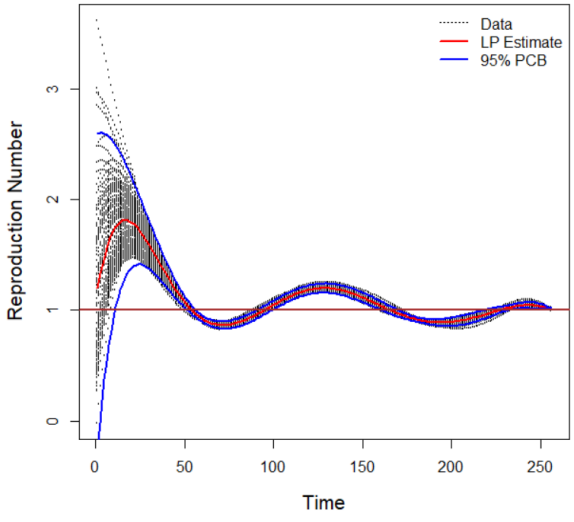

Figure 4: Georgia, USA: 95 % PCB for mean reproduction number curve

decisions during this period of time) [31]. In Figure 3, a bundle of reconstructed values of the effective reproduction number $\mathcal{R}(t)$ is illustrated. It starts off with a rather high level at the onset of the outbreak. Then, about 50 days into the process (around the time when some businesses are allowed to reopen), its value drops under 1, generating a steady flow of cases in the subsequent months. Then (seemingly right after Memorial Day weekend) $\mathcal{R}(t)$ becomes greater than 1, causing an alarming wave of cases, which starts two weeks later. At the end of the study period, incidence cases are slightly on the rise yet again, and that is consistent with the reconstructed values of the reproduction number that are just slightly greater than 1. The behavior of the reconstructed transmission rate $\beta(t)$ mimics the behavior of the reproduction number $\mathcal{R}(t)$ to a large extent (Figure 5, left).

The calculated 95 % pointwise confidence bands (PCB) for the effective reproduction number $\mathcal{R}(t)$ (Figure 4) show extremely narrow confidence intervals for the most part of the study period, which indicate that the algorithm we used to extract the parameters from full data sets produced highly consistent results. As

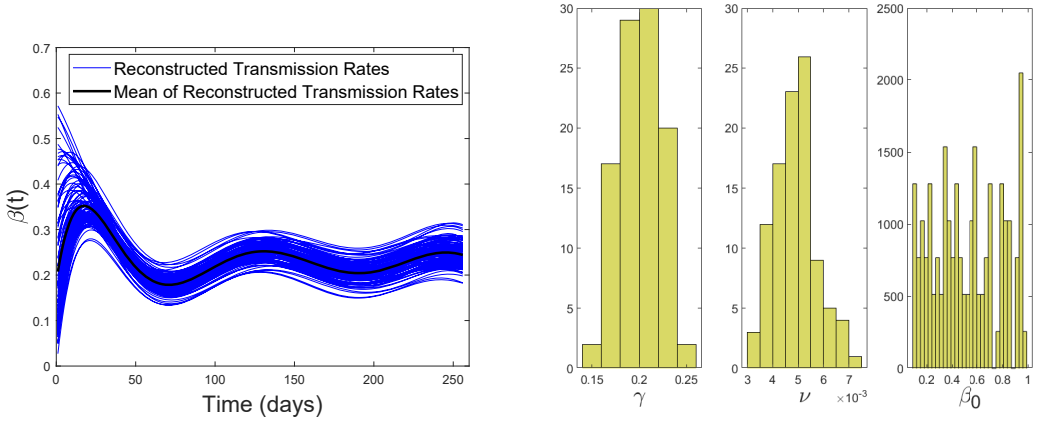

Figure 5: Reconstructed transmission rate: Georgia (left) and simulated values of γ, ν, and β_0 (right)

expected, the bands are wider in the beginning of the outbreak, but they narrow down as time goes on. This provides convincing evidence that the forecasting methods applied to this model have the potential to generate accurate predictions. The reconstructed value of the reporting rate ψ for the state of Georgia is 0.23 (95 % CI: $[0.22, 0.24]$), which does not come as a surprise considering a large number of mild and asymptomatic cases.

In our experiments for the state of Georgia, for every bootstrap iteration, we take $\alpha_0 = 10^{-3}$ in (3.2), but a wide range of values from $\alpha_0 = 10^{-1}$ to $\alpha_0 = 10^{-8}$ can be used to get the results that are almost identical. The convergence rate for $\{\alpha_k\}$ is chosen to be $\alpha_k = \alpha_0/k$, the rate that gives rise to the most stable iterative process. Iterations are terminated after 10 predictor-corrector steps. Due to very sporadic nature of reported data on COVID-19 incidence cases and deaths, we discretize $\frac{dI}{dt}(t)$ and $\beta(t)$ with 30 and 8 base functions, respectively (that is, $N = 15$ and $m = 8$). The population of the state of Georgia was reported at 10.62 million in 2019 [33].

For the second numerical experiment, we take data on incidence cases and deaths for the COVID-19 virus in the state of New York, USA, over the period of 255 days, from March 1, 2020, to November 10, 2020 (see Figures 6 and 7 for the data used and for the major public health decisions during this period of time) [32].

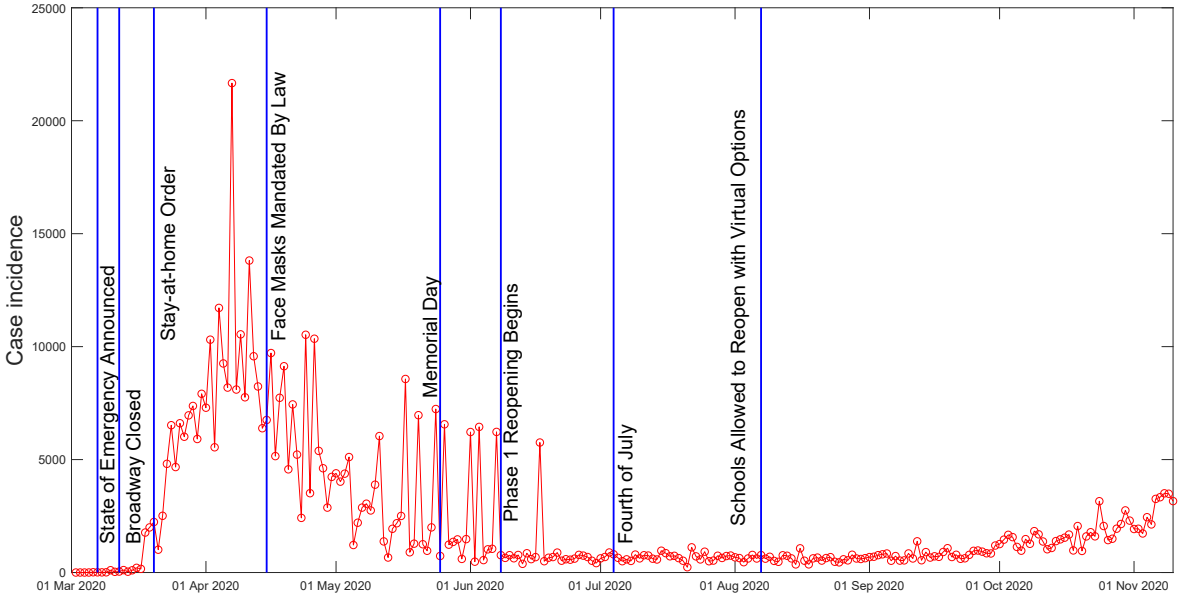

Figure 6: The state of New York, USA, COVID-19 incidence wave, 2020

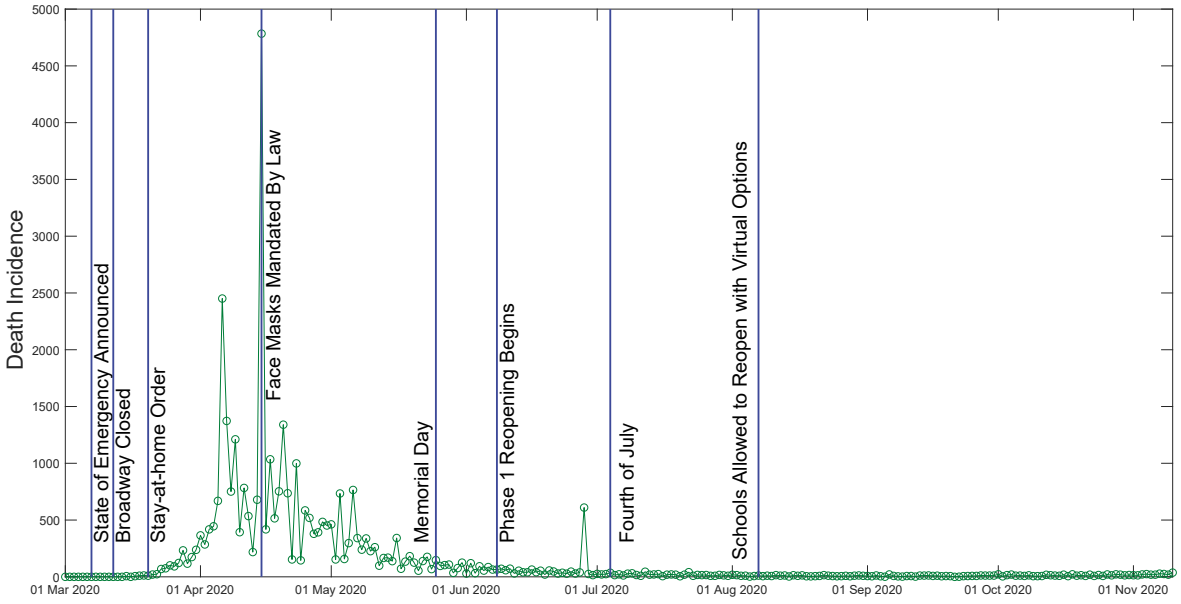

Figure 7: The state of New York, USA, COVID-19 death wave, 2020

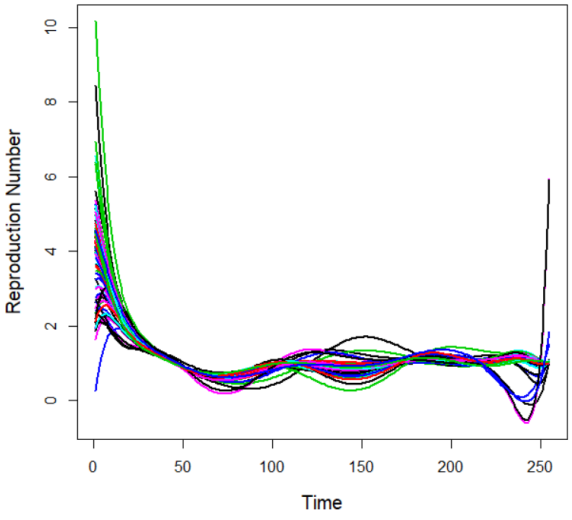

Figure 8: New York, USA: effective reproduction number as functional data

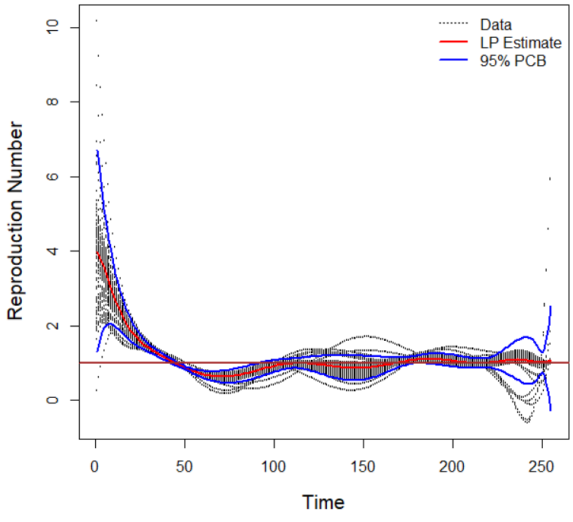

Figure 9: New York, USA: 95 % PCB for mean reproduction number curve

In Figure 8, a bundle of reconstructed values of the effective reproduction number $\mathcal{R}(t)$ for New York is shown. From the graph of the bundle, we can see that the behavior of the reproduction curves varies greatly from those in Georgia. The dramatic increase in incidence cases in the state of New York at the onset of the virus justifies high values of the reproduction number in the beginning of the study window, resulting in wider confidence bands at this time, as illustrated in Figure 9. After the initial spike, 95 % pointwise confidence bands become rather narrow (but not quite as narrow as in the case of Georgia) indicating that our algorithm produces consistent and reliable results. The bands widen yet again near the 120 day mark (and until day 175). Evidently, this is the result of the corresponding COVID-19 data being rather chaotic. After this period of time, we see the bands narrow for the next 50 days showing consistent confidence in the reproduction mean value.

Overall, it appears that, with strict social distancing measures, the effective reproduction number $\mathcal{R}(t)$ drops under 1 after about 45 days into the outbreak, producing the inflection point in the incidence curve

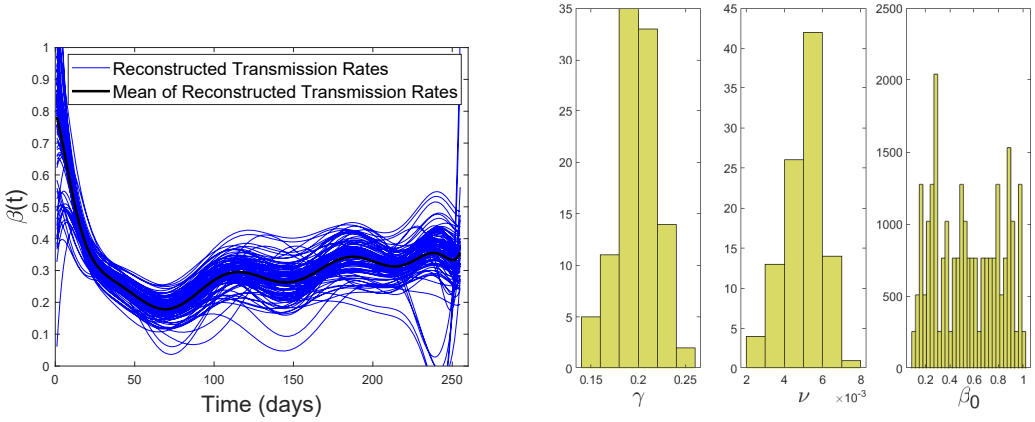

Figure 10: Reconstructed transmission rate: New York (left) and simulated values of γ, ν, and β_0 (right)

about two weeks later followed by a steady (and relatively low) flow of cases up until the end of the interval. There is, however, a disturbing uphill trend in the incidence curve towards the end of the study period, which causes uncertain behavior in the reconstructed reproduction number. For the reconstructed transmission rate in the state of New York, the uphill trend over the last 100 days of the study period is even more pronounced (Figure 10, left).

The good news is that this trend is not observed in the reported data on new deaths, indicating that, unlike our assumption in (1.1)–(1.4), the death rate does not actually remain a constant: it goes down. While experimenting with New York data sets, for every bootstrap iteration, we take $\alpha_0 = 10^{-2}$, though a wide range of values from $\alpha_0 = 10^{-1}$ to $\alpha_0 = 10^{-5}$ can be used to get very similar results. The convergence rate for $\{\alpha_k\}$ is chosen to be $\alpha_k = \alpha_0/k$, the rate that gives rise to the most rapid (yet stable) iterative process. Since the New York data is extremely spread out, we discretize $\frac{dI}{dt}(t)$ and $\beta(t)$ with 16 and 12 base functions, respectively (that is, $N = 8$ and $m = 12$). The population of the state of New York was reported at 19.45 million as of July 2019 [34].

To address the problem of semi-convergence of computational algorithms in the ill-posed case [1, 3, 14, 26], the iterative process has been terminated after 15 predictor-corrector steps to avoid over-fitting. According to our model, the recovered value of the reporting rate ψ for New York is 0.063 (95 % CI: [0.058, 0.067]), which is much less and much more uncertain as compared to Georgia.

Upon comparing different choices of the penalty operator T in (3.2), it has been concluded that $T^*T = \mathcal{I}$, the identity operator in $\mathbb{R}^m$, gives rise to the most stable reconstruction of the unknown parameter θ and the unknown state variable u. To ensure that the cost functional (3.4) is equally sensitive to $d_\delta^{(1)}$ and $d_\delta^{(2)}$, we take $\lambda_1 = 1$ and $\lambda_2 = 500$, and $\lambda_1 = 1$ and $\lambda_2 = 150$ in iterative scheme (3.3) for Georgia and New York, respectively.

Figures 5 (Georgia) and 10 (New York) illustrate reconstructed transmission rates $\beta(t)$ (left) and the simulated values of the recovery rate γ, fatality rate ν, and initial approximation for the transmission rate β_0 (right). To study the sensitivity of our model to inevitable variations in γ and ν, within each of the 100 bootstrap iterations, we sample γ and ν from normal distributions, $N(0.20, 0.02)$ and $N(0.005, 0.001)$, respectively, while using uniform distribution on $[0.1, 1]$ for β_0. We let γ follow the normal distribution, $N(0.20, 0.02)$, to reflect an average infectious period between 3 and 20 days. We assume the mean value of the fatality rate ν to be 0.5 % [2]. The fatality rate ν follows the normal distribution $N(0.005, 0.001)$ to account for variation of this parameter within different risk groups. The reconstructed values of $\beta(t)$ with normally distributed γ and ν are virtually identical to those reconstructed with constant (mean) values of γ and ν, which highlights a very low sensitivity of $\beta(t)$ to variations in COVID-19 infectious periods and fatality rates in individuals contracting this disease. The reconstructed values of $\beta(t)$ are also quite immune to variations of $\beta_0 \in [0.1, 1]$, which shows that all these values of β_0 are within the radius of convergence of our parameter estimation algorithm.

As emphasized above, the upward trend in the New York transmission rate over the last three months of the study period is concerning. And for both states, even though the values of the transmission rate are much lower towards the end of the interval (as opposed to what we see at the start of the pandemic), the estimated transmission rates clearly indicate that the outbreak is far from over at this time.

7 Forecasting strategy and results

When epidemiological parameters in a mathematical model of an infectious disease are assumed to be constant, the forecasting strategy of future incidence cases and/or new daily deaths is straightforward: one substitutes the values of the parameters, recovered during the calibration period, into the differential equation (or the system of differential equations) and solves this equation/system for future values of time [22]. However, it has been established in multiple studies [5, 19] that a transmission rate of a virus (and the resulting value of the effective reproduction number) is directly impacted by intervention and control (as well as numerous other factors) and, as such, it is time dependent.

The expansion coefficients for its Legendre (or other) approximation are only accurate over the interval of time where the epidemiological data is available. Moving forward, the Legendre approximation of the disease transmission rate $\beta(t)$ is completely random and cannot be used to generate future projections of new incidence cases. The same is true for Fourier expansions of the state variables in our disease model. Hence, for future values of time, one needs to evaluate some parametric representation of epidemiological parameters that would, in a meaningful way, extrapolate the behavior observed during the calibration period. In this section, to forecast future COVID-19 cases, we propose parametric extrapolation of $\mathcal{R}(t)$ based on Weibull functions [2].

Assume that we have complete data on incidence cases and daily deaths, $d^{(1)}$ and $d^{(2)}$, available for some region from days t_1 to T, where t_1 is the day when the first COVID-19 case is reported and T represents the current time. To forecast future incidence cases (for $t > T$), we propose the following strategy. First, for $t \in [t_1, T]$, we use predictor-corrector algorithm (3.2)–(3.3), along with Fourier approximation for the unknown state variables and the approximation by Legendre polynomials for the unknown transmission rate, in order to estimate $\beta(t)$, constrained by (1.1)–(1.4), and then $\mathcal{R}(t)$, defined in (1.5). Second, on the interval $[T - \tau, T]$, $\tau < T - t_1$, we approximate the effective reproduction number $\mathcal{R}(t)$ by a parametric function $\tilde{\mathcal{R}}(t)$ that is based on Weibull specification with one regime as suggested in [2],

$$\tilde{\mathcal{R}}(t) = 1 + \frac{1}{\gamma}\Big[\frac{\mathcal{B}-1}{t-t_1-\mathcal{C}} - \frac{\mathcal{B}}{\mathcal{A}}\Big(\frac{t-t_1-\mathcal{C}}{\mathcal{A}}\Big)^{\mathcal{B}-1}\Big], \tag{7.1}$$

where $\mathcal{A} > 0$ is the scale parameter, $\mathcal{B} > 0$ is the slope parameter, and $\mathcal{C}$ is the location parameter of Weibull density. Among various other parametric representations tried by the authors, Weibull specification (7.1) proved to be the most efficient. It allows to accurately capture the dynamic of the estimated reproduction number $\mathcal{R}(t)$ during the calibration (training) period $[T - \tau, T]$ and to extrapolate it to the forecasting period $[T, T + \varepsilon]$. Parametrization (7.1) does not oversimplify the effective reproduction number $\mathcal{R}(t)$. Yet, it does not add any extra features to the extrapolated values of $\mathcal{R}(t)$, which may cause the change in behavior that is not justified and that can potentially result in inaccurate forecasting.

In order to fit the function $\tilde{\mathcal{R}}(t)$, defined in (7.1), to the estimated values of the effective reproduction number $\mathcal{R}(t)$, we use the Matlab built-in solver *lsqcurvefit* that implements a regularized Levenberg–Marquardt algorithm for nonlinear optimization [15]. To enforce stability, we introduce the following notations:

$$\xi := t_1 + \mathcal{C}, \quad \zeta := \frac{\mathcal{B}}{\mathcal{A}^{\mathcal{B}}} \quad \text{and} \quad \eta := \mathcal{B} - 1,$$

which yield the modified identity

$$\tilde{\mathcal{R}}(t) = 1 + \frac{1}{\gamma}\Big[\frac{\eta}{t-\xi} - \zeta(t-\xi)^{\eta}\Big]. \tag{7.2}$$

According to (1.2) and (1.5), one has

$$\frac{dI}{dt} = (\mathcal{R}(t) - 1)\gamma I(t).$$

Let $I_1 = I_1(t)$ be our estimation of $I = I(t)$ by predictor-corrector algorithm (3.2)–(3.3) on the interval $[t_1, T]$, and let $I_2 = I_2(t)$ be the projected values of $I = I(t)$ on the interval $[T, T + \varepsilon]$. To find $I_2 = I_2(t)$, once parameters ξ, ζ, and η in (7.2) have been calculated from the function $\mathcal{R}(t)$, one solves the differential equation $\frac{dI}{dt} = (\tilde{\mathcal{R}}(t) - 1)\gamma I(t)$. Thus,

$$I_2(t) = I_1(T)\Big[\frac{t-\xi}{T-\xi}\Big]^{\eta} \exp\Big[-\frac{\zeta}{\eta+1}\{(t-\xi)^{\eta+1} - (T-\xi)^{\eta+1}\}\Big], \quad t \in [T, T+\varepsilon]. \tag{7.3}$$

This analytic solution agrees with the numerical solution, obtained by the built-in solver ode23s of Matlab. Given (7.3), one can project future incidence cases and daily deaths using the expressions

$$\text{new incidence cases} \approx \psi\Big(\frac{dI_2}{dt} + \gamma I_2\Big) \approx \frac{\big(\frac{dI_1}{dt} + \gamma I_1, d^{(1)}\big)}{\big\|\big(\frac{dI_1}{dt} + \gamma I_1\big)\big\|^2}\tilde{\mathcal{R}}\,\gamma\,\tilde{I}_2,$$

$$\text{daily deaths} \approx \nu\gamma\tilde{I}_2,$$

respectively. In the above, the scalar product, $(\cdot,\cdot)$, and the Euclidian norm, $\|\cdot\|$, are for the partition of the interval $[t_1, T]$, and ψ is the reporting rate for new incidence cases, estimated on the interval $[t_1, T]$, where the data $d^{(1)}$ is available.

In our first forecasting experiment, we use 60 days of data on new incidence cases and daily deaths in the state of Georgia, USA, in order to forecast 30 days forward. The unknown parameters ζ, η, and ξ in the expression for $\tilde{\mathcal{R}}(t)$ are estimated over a 1-week training period (from day 53 to day 60), while the unknown reporting rate ψ is estimated based on all available data from day 1 to day 60. The reconstructed values of ζ, η, ξ, and ψ are illustrated in Table 1.

For every bootstrap iteration, we take $\alpha_0 = 10^{-3}$, but a wide range of values from $\alpha_0 = 10^{-1}$ to $\alpha_0 = 10^{-9}$ can be used to get the results that are almost indistinguishable. The convergence rate for $\{\alpha_k\}$ is chosen to be $\alpha_k = \alpha_0/k$, the most aggressive convergence rate that maintains stability until the iterative process is termi-

$N = 5$, $m = 6$, $\lambda_1 = 1$, $\lambda_2 = 1$, $\alpha_k = \alpha_0/k$, and $\alpha_0 = 10^{-3}$	
$\zeta = -0.098$ (95 % CI: $[-0.12, 0.00023]$)	$\eta = -0.44$ (95 % CI: $[-0.57, 0.0036]$)
$\xi = 45$ (95 % CI: $[35, 52]$)	$\psi = 0.14$ (95 % CI: $[0.088, 0.26]$)

Table 1: Georgia, USA: 95 % confidence intervals for reconstructed parameters with 1-week training period for $\mathcal{R}(t)$ (from day 53 to day 60)

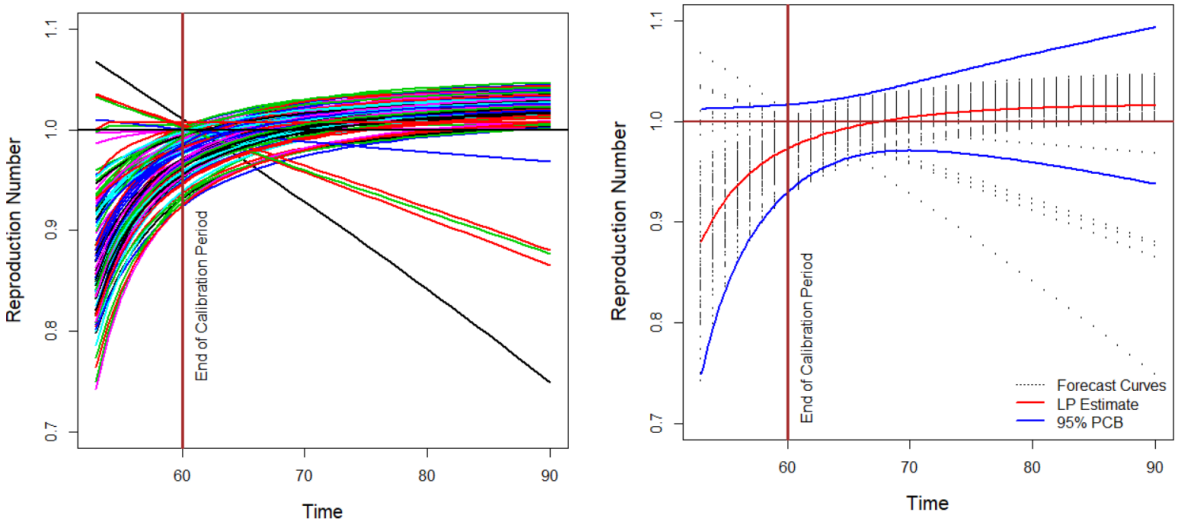

Figure 11: Georgia, USA: $\mathcal{R}(t)$ projected 30 days forward with 1-week training period from day 53 to day 60 as functional data (left) and 95 % PCB (right)

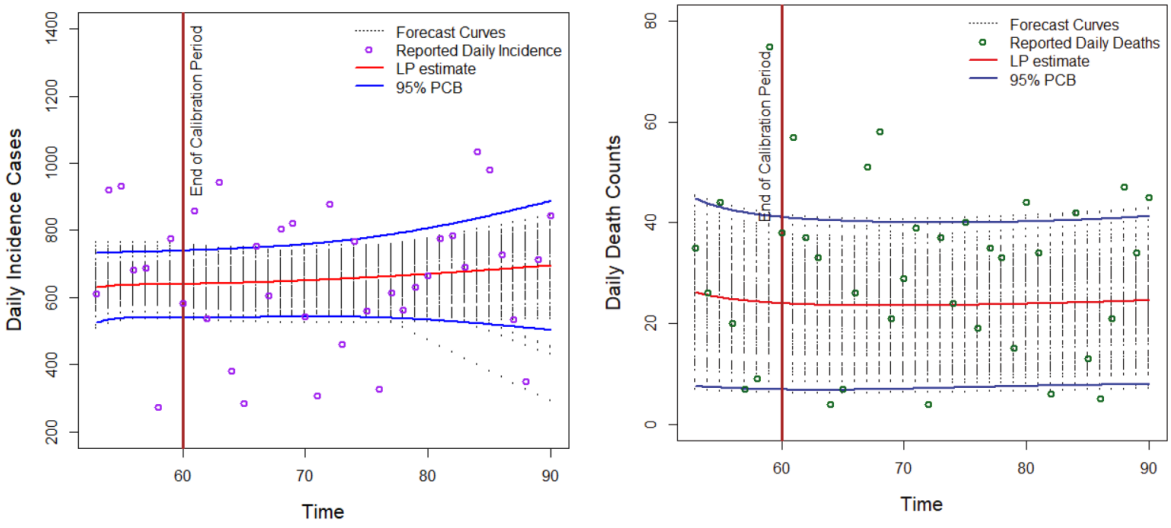

Figure 12: Georgia, USA: incidence cases (left) and daily deaths (right) projected 30 days forward with 1-week training period for $\mathcal{R}(t)$ (from day 53 to day 60)

$N = 4$, $m = 10$, $\lambda_1 = 1$, $\lambda_2 = 1$, $\alpha_k = \alpha_0/k$, and $\alpha_0 = 10^{-3}$	
$\zeta = 0.00088$ (95 % CI: $[1.7e-05, 0.0048]$)	$\eta = 1.1$ (95 % CI: $[0.43, 1.9]$)
$\xi = 1.1e+02$ (95 % CI: $[92, 1.3e+02]$)	$\psi = 0.28$ (95 % CI: $[0.24, 0.32]$)

Table 2: Georgia, USA: 95 % confidence intervals for reconstructed parameters with 1-week training period for $\mathcal{R}(t)$ (from day 143 to day 150)

nated. The number of predictor-corrector steps in each bootstrapping loop is equal to 10. We discretize $\frac{dI_1}{dt}(t)$ and $\beta(t)$ with 10 and 6 base functions, respectively (that is, $N = 5$ and $m = 6$).

As evident from Figure 11, the extrapolated values of the effective reproduction number $\mathcal{R}(t)$ suggest that the reproduction number will be on the rise for the duration of the forecasting period (between days 60 and 90), which is consistent with the upward trend seen in Figure 4, where $\mathcal{R}(t)$ is estimated from full data. However, the actual mean values of $\tilde{\mathcal{R}}(t)$ are higher as compared to the corresponding values of $\mathcal{R}(t)$ between days 60 and 90, presented in Figure 4. Still, the forecasting bundle for projected new incidence cases and daily deaths covers the reported data quite well as shown in Figure 12.

In our second forecasting experiment, we use 150 days of Georgia data in order to generate a 30-day forecast. The unknown parameters ζ, η, and ξ in the expression for $\tilde{\mathcal{R}}(t)$ are estimated over a 1-week training period (from day 143 to day 150), and the unknown reporting rate for new incidence cases ψ is estimated based on all available data from day 1 to day 150. The reconstructed values of ζ, η, ξ, and ψ for the second data set are illustrated in Table 2.

For every bootstrap iteration, we take $\alpha_0 = 10^{-3}$. As in the case of the previous experiment, values from $\alpha_0 = 10^{-1}$ to $\alpha_0 = 10^{-9}$ can also be used. The regularization sequence $\{\alpha_k\}$ converges at the rate $\alpha_k = \alpha_0/k$, and the number of predictor-corrector steps in each bootstrapping loop is equal to 10. We discretize $\frac{dI_1}{dt}(t)$ and $\beta(t)$ with 8 and 10 base functions, respectively (that is, $N = 4$ and $m = 10$), and we take $\lambda_1 = \lambda_2 = 1$. Note that the range of acceptable parameters is sufficiently broad, and simulations with, for example, $N = 6$ and $m = 12$, $\lambda_1 = 1$, and $\lambda_2 = 2$ generate forecasting bundles that are virtually identical. From Table 2, one can see that, as we use more data, the value of ψ, the reporting rate for new incidence cases, is getting closer to the reporting rate recovered from full Georgia data set (and the confidence interval is getting narrower). As stated in Section 6, for the full data set, $\psi = 0.23$ (95 % CI: $[0.22, 0.24]$).

As shown in Figure 13, at the end of the calibration period, new incidence cases and daily deaths begin to slide down. This trend is adequately captured and carried into the forecasting interval (from day 150 to day 180), with both forecasting bundles decreasing at the rate that is consistent with reported data. As com-

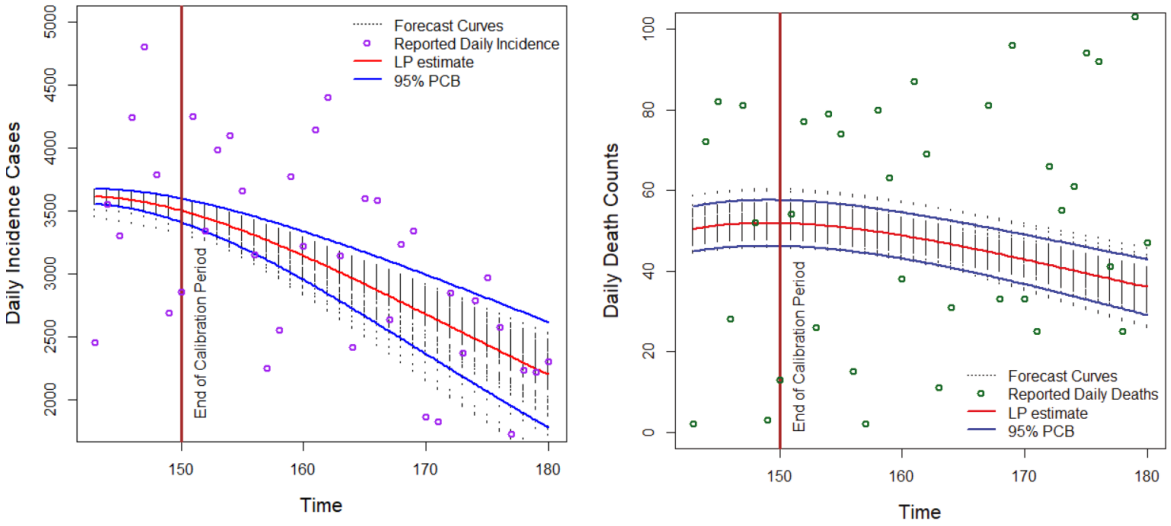

Figure 13: Georgia, USA: incidence cases (left) and daily deaths (right) projected 30 days forward with 1-week training period for $\mathcal{R}(t)$ (from day 143 to day 150)

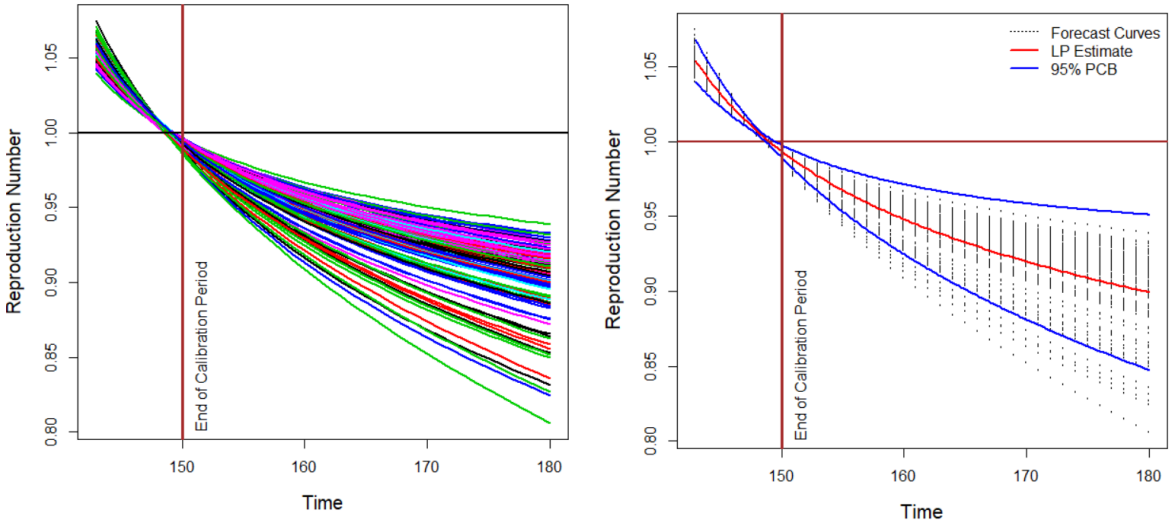

Figure 14: Georgia, USA: $\mathcal{R}(t)$ projected 30 days forward with 1-week training period from day 143 to day 150 as functional data (left) and 95 % PCB (right)

pared to the reported data, the estimated mean values for new cases and deaths are accurate, with narrow confidence intervals. The corresponding values of the projected reproduction number are presented in Figure 14.

Next, we conduct numerical simulations with partial data for the state of New York, USA. Initially, we use only 20 days of data in order to forecast 30 days forward. To capture the most recent trend in the behavior of the reproduction number and to extrapolate it further, the unknown parameters ζ, η, and ξ in (7.2) are estimated over a 1-week training period from day 23 to 30, while the unknown reporting rate for new incidence cases ψ is estimated based on all available data from day 1 to day 20. The reconstructed values of ζ, η, ξ, and ψ for this data set are illustrated in Table 3.

For all 100 bootstrap iterations, we take $\alpha_0 = 10^{-5}$. The sequence $\{\alpha_k\}$ goes to zero at the rate α_0/k, and each predictor-corrector step is terminated after 10 iterations. To reduce noise propagation, we discretize $\frac{dI_1}{dt}(t)$ with 6 Fourier functions (i.e., $N = 3$), while $\beta(t)$ is approximated by a linear combination of 8 Legendre polynomials ($m = 8$). Given small values of $d^{(2)}$ during the calibration period, we put a little more weight on the third term in the cost functional (3.4) by setting $\lambda_2 = 2$ and $\lambda_1 = 1$.

$N = 3$, $m = 8$, $\lambda_1 = 1$, $\lambda_2 = 2$, $\alpha_k = \alpha_0/k$, and $\alpha_0 = 10^{-5}$	
$\zeta = -0.17$ (95 % CI: $[-0.19, -0.16]$)	$\eta = -0.023$ (95 % CI: $[-0.034, -0.017]$)
$\xi = 17$ (95 % CI: $[17, 17]$)	$\psi = 0.9$ (95 % CI: $[0.54, 1.3]$)

Table 3: New York, USA: 95 % confidence intervals for reconstructed parameters with 1-week training period for $\mathcal{R}(t)$ (from day 13 to day 20)

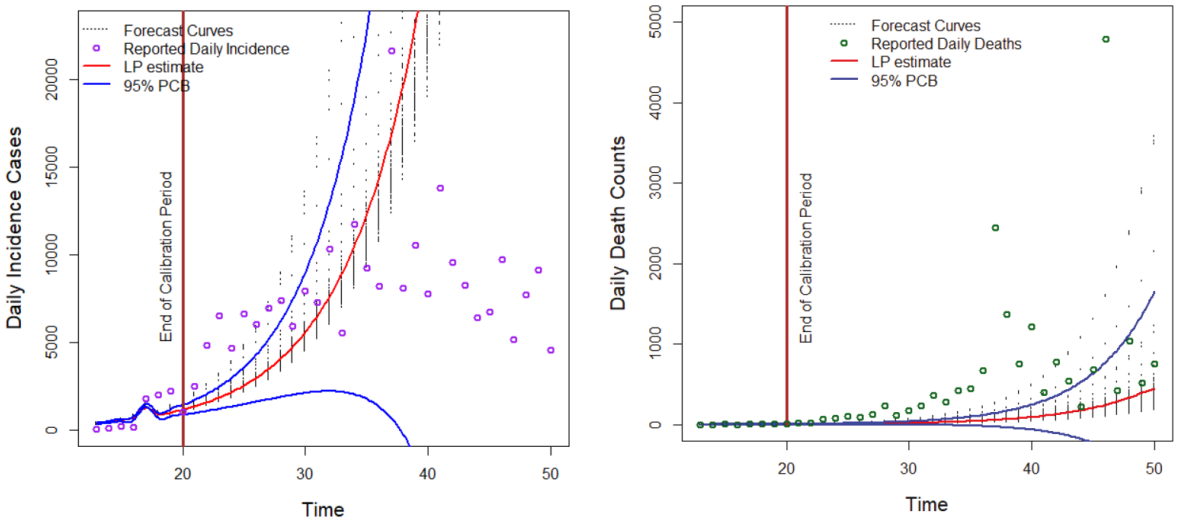

Figure 15: New York, USA: incidence cases (left) and daily deaths (right) projected 30 days forward with 7-day training period for $\mathcal{R}(t)$ (from day 13 to day 20)

With very few new deaths reported daily between day 1 and day 20, the algorithm fails to accurately estimate the reporting rate for incidence cases ψ. While full data for the state of New York suggest that, on average, only 6.3 % of new COVID-19 cases get reported (that is, $\psi = 0.063$ (95 % CI: $[0.058, 0.067]$), the approximate value of ψ, calculated from 20 data points, is 0.9 (95 % CI: $[0.54, 1.3]$). This results in substantial overestimation of new incidence cases during the second half of the forecasting period (Figure 15). Yet, in the first 15 days, the projected values for new incidence cases are accurate. For daily deaths, the forecasting curves underestimate the data for the entire 30 day interval. They do not capture a spike that occurs between 35 and 40 days, nor do they follow a huge jump that happens on day 46. On the bright side, the mean value for daily deaths does predict the uphill trend towards the end of the study period, even though the projected values are slightly lower than the actual data.

Another reason for inaccurate incidence forecasting (apart from the overestimated value of ψ) is a considerable variation in the behavior of the reproduction number during the training time as shown in Figure 16. That makes it difficult to predict how the behavior will change in the upcoming 30 days. And, of course, the current model does not allow to account for any changes in control policies *after* the calibration period. So, in a way, the forecasting mean only predicts what will happen under current mitigation measures. If, however, these measures become more efficient, then the actual number of new cases is lower than its projected value. Hence, the best forecasting strategy is to recalculate projected values every week (or 10 days) as new data become available.

In our last experiment, for the sake of forecasting 1 month forward, we use 130 days of data on new incidence cases and daily deaths in the state of New York, USA. The unknown parameters ζ, η, and ξ for the projected reproduction number $\tilde{\mathcal{R}}(t)$ are estimated over a 1-week training period (from day 123 to day 130), while the unknown reporting rate ψ is estimated based on all available data for 130 days. The reconstructed values of ζ, η, ξ, and ψ are given in Table 4.

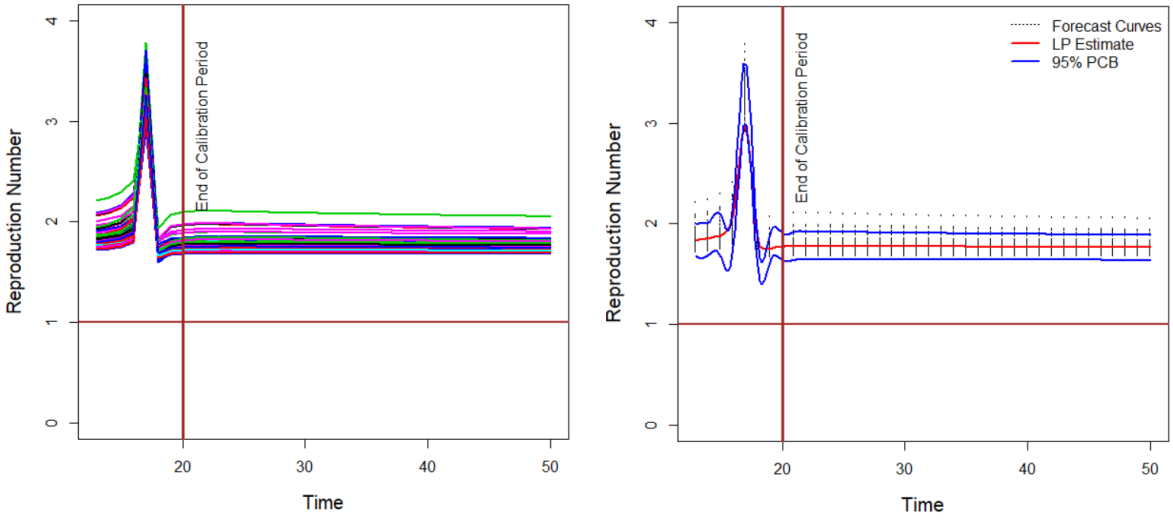

Figure 16: New York, USA: $\mathcal{R}(t)$ projected 30 days forward with 7-day training period from day 13 to day 20 as functional data (left) and 95 % SCB (right)

For every bootstrap iteration, we take $\alpha_0 = 10^{-5}$ with $\alpha_k = \alpha_0/k$ in order to carry out 10 predictor-corrector steps. We discretize $\frac{dI_1}{dt}(t)$ and $\beta(t)$ with 8 and 9 base functions, respectively (that is, $N = 4$ and $m = 9$). As one can see in Figure 17, during the forecasting period, the effective reproduction number is projected to remain nearly flat with its values close to the critical threshold level $\tilde{\mathcal{R}}(t) = 1$. Given this pattern, $\tilde{\mathcal{R}}(t)$ is expected to generate a steady flow of new incidence cases and daily deaths, which is fully consistent with the reported data. Figure 18 (left) shows that the forecasting bundle covers most of new incidence data

$N = 4$, $m = 9$, $\lambda_1 = 1$, $\lambda_2 = 1$, $\alpha_k = \alpha_0/k$, and $\alpha_0 = 10^{-5}$	
$\zeta = -0.12$ (95 % CI: $[-0.13, -0.089]$)	$\eta = -0.55$ (95 % CI: $[-0.63, -0.52]$)
$\xi = 1.2e+02$ (95 % CI: $[1.1e+02, 1.2e+02]$)	$\psi = 0.19$ (95 % CI: $[0.16, 0.23]$)

Table 4: New York, USA: 95 % confidence intervals for reconstructed parameters with 1-week training period for $\mathcal{R}(t)$ (from day 123 to day 130)

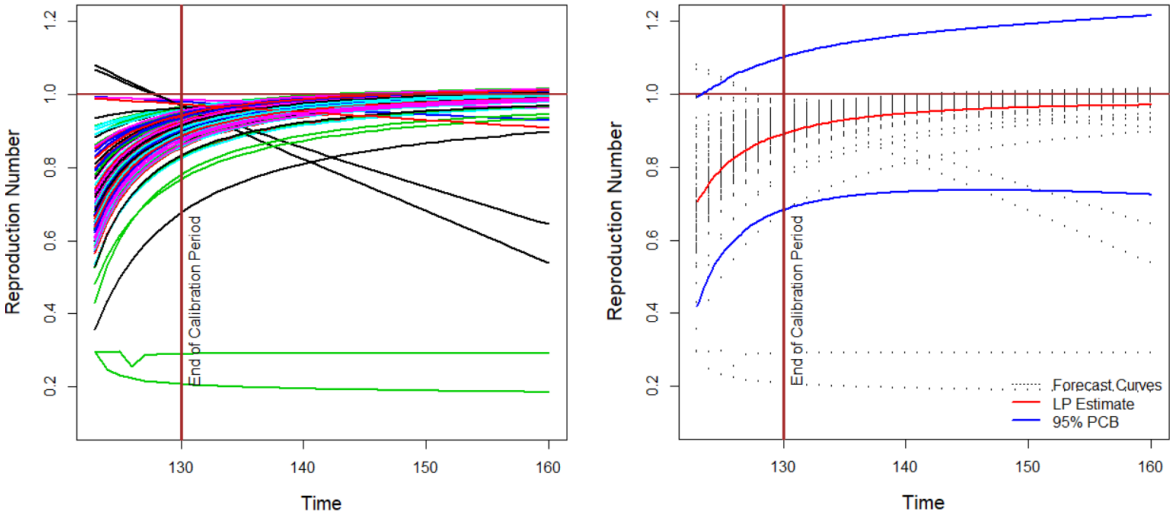

Figure 17: New York, USA: $\mathcal{R}(t)$ projected 30 days forward with 7-day training period from day 123 to day 130 as functional data (left) and 95 % PfCB (right)

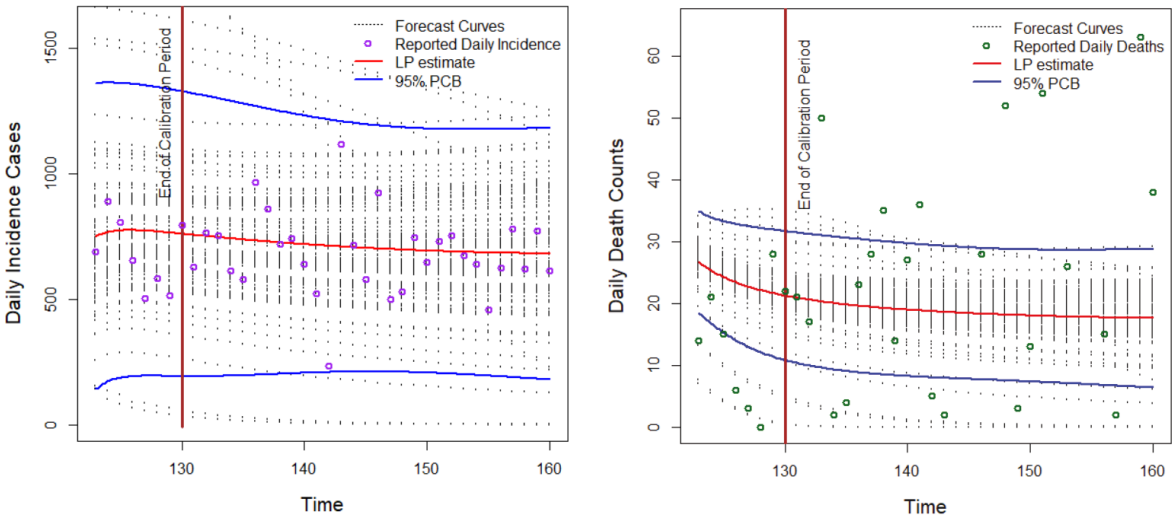

Figure 18: New York, USA: incidence cases (left) and daily deaths (right) projected 30 days forward with 7-day training period for $\mathcal{R}(t)$ (from day 123 to day 130)

points between days 130 and 160, and Figure 17 (right) illustrates a rather good match of the corresponding projection curves to the daily deaths data. The reporting rate ψ, calculated from 130 data points, is getting closer to the one estimated from the full data set (0.063 (95 % CI: [0.058, 0.067]).

8 Conclusions and future plans

In this paper, we introduce a novel optimization algorithm for stable parameter estimation and forecasting of future incidence cases for COVID-19 outbreak in the states of Georgia and New York, USA. All experiments are carried out with real data on the COVID-19 pandemic in the USA. The algorithm combines a version of the SIRD compartmental model of disease progression, proposed in [2], with iteratively regularized predictor-corrector numerical scheme [22], aimed at the reconstruction of COVID-19 reporting ratio ψ, transmission rate $\beta(t)$, and effective reproduction number $\mathcal{R}(t)$. Our creative forecasting approach uses the idea of Weibull functions [2] in order to approximate $\mathcal{R}(t)$, which allows to closely follow the dynamic of the computed reproduction number during the calibration time $[T-\tau, T]$ and then to extrapolate it to the forecasting period $[T, T+\varepsilon]$.

It is evident that our ability to quantify the reporting rate for COVID-19 daily incidence cases and deaths has made it possible to improve the accuracy of future projections and to better assess the efficiency of government intervention measures during various phases of the ongoing pandemic. At the same time, our numerical simulations have shown that a constant death rate ν is an oversimplification, which, at times, resulted in a mismatch between predicted values of future new incidence cases and daily deaths. Moving forward, apart from the terms accounting for the availability of the vaccine, a more realistic parametric representation for COVID-19 death rate needs to be incorporated in the SIRD model. For uncertainty quantification, we plan to explore the possibility of using the functional data analysis to compute 95 % simultaneous (uniform) confidence bands for the mean value of $\mathcal{R}(t)$, i.e., such confidence bands that the true curve falls entirely within the band 95 % of the time (as opposed to pointwise confidence intervals that have been constructed in this paper).

Funding: A. Smirnova was supported by NSF awards 1818886 and 2011622 (DMS Computational Mathematics).

References

[1] R. C. Aster, B. Borchers and C. H. Thurber, *Parameter Estimation and Inverse Problems*, Academic Press, New York, 2011.

[2] A. Atkeson, K. Kopecky and T. Zha, Estimating and Forecasting Disease Scenarios for COVID-19 with an SIR Model, NBER Working Paper No. 27335, 2020, NBER Program(s): Economic Fluctuations and Growth.

[3] A. B. Bakushinsky and M. Y. Kokurin, *Iterative Methods for Ill-Posed Operator Equations with Smooth Operators*, Springer, Dordrecht, 2004.

[4] G. Chowell, Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts, *Infect. Dis. Model.* **2** (2019), 379–398.

[5] G. Chowell and H. Nishiura, Transmission dynamics and control of Ebola virus disease (EVD): A review, *BMC Med.* **12** (2014), DOI 10.1186/s12916-014-0196-0.

[6] G. Chowell, A. Tariq and J. M. Hyman, A novel sub-epidemic modeling framework for short-term forecasting epidemic waves, *BMC Med.* **17** (2019), Article ID 164.

[7] B. Efron and R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statist. Sci.* **1** (1986), 54–75.

[8] H. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic, Dordecht, 1996.

[9] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo and M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, *Nat. Med.* **26** (2020), 855–860.

[10] Q. Jin and W. Wang, Analysis of the iteratively regularized Gauss–Newton method under a heuristic rule, *Inverse Problems* **34** (2018), no. 3, Article ID 035001.

[11] B. Kaltenbacher, A. Neubauer and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, Radon Ser. Comput. Appl. Math. 6, Walter de Gruyter, Berlin, 2008.

[12] J. O. Lloyd-Smith, S. Funk, A. R. McLean, S. Riley and J. L. Wood, Nine challenges in modelling the emergence of novel pathogens, *Epidemics* **10** (2015), 35–39.

[13] S. Morse, J. Mazet, M. Woolhouse, C. Parrish, D. Carroll, W. Karesh and P. Daszak, Prediction and prevention of the next pandemic zoonosis, *The Lancet* **380** (2012), 1956–1965.

[14] A. Neubauer, Optimal convergence rates for inexact Newton regularization with CG as inner iteration, *J. Inverse Ill-Posed Probl.* **28** (2020), 145–153.

[15] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, Cham, 2000.

[16] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 2014.

[17] J. O. Ramsay, S. Graves and G. Hooker, FDA: Functional Data Analysis. R package version 5.1.4., 2020, https://CRAN.R-project.org/package=fda.

[18] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, Cham, 2005.

[19] E. O. Romero-Severson, N. Hengartner, G. Meadors and R. Ke, Change in global transmission rates of COVID-19 through May 6, *PLoS ONE* **15** (2020), no. 8, Article ID e0236776.

[20] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. M. Hyman, P. Yan and G. Chowell, Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, *J. Clin. Med.* **9** (2020), Article ID 596.

[21] A. Smirnova, On convergence rates for iteratively regularized procedures with a linear penalty term, *Inverse Problems* **28** (2012), no. 8, Article ID 085005.

[22] A. Smirnova and A. Bakushinsky, On iteratively regularized predictor-corrector algorithm for parameter identification, *Inverse Problems* **36** (2020), Article ID 125015.

[23] A. Smirnova, G. Chowell, L. DeCamp, S. Moghadas and M. Sheppard, Improving epidemic size prediction through stable reconstruction of disease parameters by reduced iteratively regularized Gauss–Newton algorithm, *J. Inverse Ill-Posed Probl.* **25** (2017), no. 5, 653–668.

[24] A. Smirnova, R. Renaut and T. Khan, Convergence and application of a modified iteratively regularized Gauss–Newton algorithm, *Inverse Problems* **23** (2007), no. 4, 1547–1563.

[25] R. N. Thompson, Epidemiological models are important tools for guiding COVID-19 interventions, *BMC Med.* **18** (2020), DOI 10.1186/s12916-020-01628-4.

[26] A. N. Tikhonov, A. Goncharsky, V. V. Stepanov and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*, Math. Appl. 328, Springer, Cham, 1995.

[27] V. V. Vasin and A. L. Ageev, *Ill-Posed Problems with A Priori Information*, VNU, Utrecht, 1995.

[28] J. Weitz and J. Dushoff, Modeling Post-death Transmission of Ebola: Challenges for Inference and Opportunities for Control, *Sci. Rep.* **5** (2015), Article ID 8751.

[29] F. Werner and B. Hofmann, Convergence analysis of (statistical) inverse problems under conditional stability estimates, *Inverse Problems* **36** (2020), Article ID 015004.

[30] CDC Coronavirus (COVID-19): Symptoms of Coronavirus, https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.

[31] Georgia Department of Public Health Daily Status Report https://dph.georgia.gov/covid-19-daily-status-report.

[32] 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-totalandratecases.
[33] U.S. Census Bureau, https://www.census.gov/quickfacts/GA.
[34] U.S. Census Bureau, https://www.census.gov/quickfacts/NY.
[35] WHO COVID-19 Global literature on coronavirus disease, https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/.
[36] WHO Coronavirus Disease (COVID-19) Dashboard, https://covid19.who.int/.