

Efficient calibration for imperfect epidemic models with applications to the analysis of COVID-19

Chih-Li Sung¹ and Ying Hung²

¹Department of Statistics and Probability, Michigan State University, East Lansing, United States

²Department of Statistics, Rutgers, the State University of New Jersey, New Brunswick, United States

Address for correspondence: Chih-Li Sung, Department of Statistics and Probability, Michigan State University, 619 Red Cedar Rd, East Lansing, MI 48824-1312, United States. Email: sungchih@msu.edu

Abstract

The estimation of unknown parameters in simulations, also known as calibration, is crucial for practical management of epidemics and prediction of pandemic risk. A simple yet widely used approach is to estimate the parameters by minimising the sum of the squared distances between actual observations and simulation outputs. It is shown in this paper that this method is inefficient, particularly when the epidemic models are developed based on certain simplifications of reality, also known as imperfect models which are commonly used in practice. To address this issue, a new estimator is introduced that is asymptotically consistent, has a smaller estimation variance than the least-squares estimator, and achieves the semiparametric efficiency. Numerical studies are performed to examine the finite sample performance. The proposed method is applied to the analysis of the COVID-19 pandemic for 20 countries based on the susceptible-exposed-infectious-recovered model with both deterministic and stochastic simulations. The estimation of the parameters, including the basic reproduction number and the average incubation period, reveal the risk of disease outbreaks in each country and provide insights to the design of public health interventions.

Keywords: basic reproduction number, compartmental models, kernel Poisson regression, semiparametric efficiency, stochastic simulations

1 Introduction

The coronavirus disease (COVID-19) pandemic has shown profound impacts on public health and the economy worldwide. The development of efficient and effective public health interventions to prevent major outbreaks and contain the pandemic relies heavily on a quantitative understanding regarding the spread of the virus, such as the transmission rate and the average incubation period. A commonly used approach in epidemiology is to estimate these quantities of interest using epidemic mathematical models, such as the susceptible-infected-recovered model, with agent-based simulations which capture complex social networks and global scale into the models (Epstein, 2009; Funk et al., 2009; Heesterbeek et al., 2015).

To estimate the parameters of interest, a widely used frequentist approach is to minimise the sum of the squared distances between the observed data and the simulation outputs, which is often referred to as the least-squares approach. See, for example, Chowell et al. (2004, 2003), Capaldi et al. (2012), Chowell (2017), Anastassopoulou et al. (2020), Bentout et al. (2020), Chen and Qiu (2020), and Giordano et al. (2020). This estimation approach is intuitive and easy to compute; however, it is shown in this paper that this method is *inefficient*, that is, its asymptotic variance is not theoretically minimal, particularly when the mathematical models associated with the simulators are built under certain assumptions or simplifications, which may not hold in reality. These

Received: March 16, 2021. Revised: June 17, 2023. Accepted: August 17, 2023

© The Royal Statistical Society 2023.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

simulators are called *imperfect* simulators in the computer experiment literature (Kennedy & O'Hagan, 2001; Plumlee, 2017; Tuo & Wu, 2015). Imperfect simulators are common in epidemiology (Heesterbeek et al., 2015), and therefore estimate parameters of interest in epidemic models based on the least-squares approach is not efficient.

To improve the estimation efficiency with imperfect epidemic models, a new estimation method is proposed in this paper. In the computer experiment literature, these unknown parameters associated with the mathematical models are often called *calibration parameters*, and the process of estimating the parameters such that the model simulations agree with the observed data is called *calibration* (Kennedy & O'Hagan, 2001; Santner et al., 2018). Although there are numerous developments on calibration, most of the work focus on continuous outputs while the discussions on non-Gaussian outputs, such as count data which are often observed in epidemiology, are scarce (Grosskopf et al., 2020; Sung, Hung, et al., 2020). In this paper, we propose a new estimation method for non-Gaussian outputs, particularly for count data for our applications in epidemiology, which minimises the L_2 projection of the discrepancy between the true mean process and the simulation outputs. It can be shown that the proposed estimator is asymptotically consistent and provides a smaller asymptotic variance than the least-squares estimator (LSE). Furthermore, it can be shown that the proposed estimator achieves the semiparametric efficiency, even when the model simulations cannot match the reality due to certain assumptions or simplifications.

It is worth noting that there are extensive studies and applications of calibration by Bayesian procedures (Diekmann et al., 2013; Farah et al., 2014; Wang, Zhou, et al., 2020; Wu et al., 2020). However, without taking the model imperfection into account in the conventional Bayesian framework, the theoretical justification for the parameter estimation with imperfect simulators are not fully developed. On the other hand, Bayesian calibration of Kennedy and O'Hagan (2001) takes into account the model imperfection through Gaussian process (GP) modelling, but it suffers from the *unidentifiability issue* when the parameter estimation is of interest (Bayarri et al., 2007; Gramacy et al., 2015; Han et al., 2009; Hodges & Riech, 2010; Paciorek, 2010). Furthermore, most of the existing developments are based on continuous outputs with a Gaussian assumption, which is not valid for the count data in the epidemic models in our applications. Recent studies on addressing the unidentifiability issue can be found in Plumlee (2017) and Tuo (2019).

The remainder of the paper is organised as follows. Two types of simulators for COVID-19 analysis, and a new estimation method based on L_2 projection for the unknown parameters in the simulators, are introduced in Section 2. Theoretical properties of the proposed estimator are developed in Section 3. In Section 4, numerical studies are conducted to demonstrate the finite sample performance of the proposed estimator and the empirical comparison with the LSE. In Section 5, the estimation method is applied to the study of COVID-19. Discussions and concluding remarks are given in Section 6. Computational details for the estimation are given in Appendix, and the mathematical proofs and the R (R Core Team, 2018) code for implementation are provided in [online supplementary material](#).

2 Estimation for compartmental models in epidemiology

2.1 Imperfect epidemic models for COVID-19 analysis

Mathematical models are commonly used in epidemiology to provide scientific insights. These models are often developed based on certain simplifications of reality; therefore, they are imperfect (Heesterbeek et al., 2015). For example, the susceptible-exposed-infectious-recovered (SEIR) model, which consists of four compartments, Susceptible-Exposed-Infectious-Recovered, is widely recommended for COVID-19 simulations because it accounts for the incubation period through the exposed compartment (Annas et al., 2020; Carcione et al., 2020; He et al., 2020; Mwalili et al., 2020; Wu et al., 2020), and is thus adopted in this paper. Mathematically, a deterministic SEIR model can be written as

$$\frac{dS}{dx} = -\frac{\beta IS}{N}, \quad \frac{dE}{dx} = \frac{\beta IS}{N} - \kappa E, \quad \frac{dI}{dx} = \kappa E - \gamma I, \quad \frac{dR}{dx} = \gamma I \quad (1)$$

where S , E , I , and R represent the numbers of cases in the corresponding compartment, $N = S + E + I + R$ is the total population, x is the time, β is the contact rate that represents the average

number of contacts per person per time in the susceptible compartment, γ is the recovery rate from the infectious compartment, and κ is the incubation rate which represents the rate of latent individuals becoming infectious, or equivalently, the average incubation period is $1/\kappa$. There are six unknown parameters in the model (1): β , κ , γ and the initial numbers of infectious, exposed, and recovered cases (denoted by $I(0)$, $E(0)$, and $R(0)$, respectively), which are denoted by $\theta = (\beta, \kappa, \gamma, I(0), E(0), R(0))$.

In this paper, we focus on two types of SEIR simulators: a deterministic simulator and a stochastic simulator. For a deterministic simulator, the simulation outputs are obtained by numerically solving the ordinary differential equations shown in (1) using numerical solvers, such as the ODEPACK (Hindmarsh, 1983). On the other hand, a stochastic SEIR simulation provides a more sophisticated and realistic framework to integrate infection dynamics in different compartments as continuous-time Markov chains (Allen, 2008, 2017; Andersson & Britton, 2012). To conduct these simulations, we implement an R package, *SimInf* (Widgren et al., 2019), in which the simulation results are obtained by the Gillespie stochastic algorithm (Gillespie, 1977). Stochastic SEIR simulations are computationally more demanding. For example, it takes more than 10 min to produce one simulation result for one country under a given parameter setting. It is computationally infeasible to perform simulations for all the possible combinations of the parameters; therefore, an *emulator* is constructed as an efficient surrogate to the actual simulation in our later implementation.

An accurate estimation of the unknown parameters in the SEIR model is often of great interest in epidemiology because it offers valuable insights into the dynamics of infectious diseases, which are essential for effectively predicting transmission patterns and assessing intervention strategies. For example, $1/\kappa$ indicates the average incubation period and the basic reproduction number, $R_0 = \beta/\gamma$, represents the expected number of new infected cases from an infectious individual in a population where all subjects are susceptible. An accurate and efficient estimation of these parameters is not only important for the public safety but it also has significant impacts on global economy. The main objective in this paper is to provide a new estimation method that enhances the estimation efficiency of parameters despite the inherent imperfections and limitations of epidemic models.

2.2 LSE and maximum likelihood estimator

Let $f(x, \theta)$ denote the number of infected cases at time $x \in \Omega \subseteq \mathbb{R}^+$, where $\theta \in \Theta \subseteq \mathbb{R}^q$ is a set of unknown calibration parameters associated with the compartmental model. In the case of SEIR model (1), $q = 6$ and $f(x, \theta) = \kappa E(x)$, where $E(x)$ is the solution of E in the ordinary differential equations of (1). Suppose that y_i is the reported number of infected cases at time x_i . Then, given the reported number of infected cases in n days, $\{(x_i, y_i)\}_{i=1}^n$, the commonly used approach to estimate the parameters is to minimise the sum of squared differences between actual numbers of infected cases and simulation outputs from compartmental models. The estimated parameters are denoted by $\hat{\theta}_n^{\text{LS}}$, where LS stands for least-squares, and they are obtained by

$$\hat{\theta}_n^{\text{LS}} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (y_i - f(x_i, \theta))^2 \quad (2)$$

In addition to the LSE, the maximum likelihood estimator (MLE) is also a commonly used estimation approach. Assume that $y_i \sim \text{Poi}(f(x_i, \theta))$, where $i = 1, \dots, n$, we obtain the MLE for the calibration parameters by

$$\hat{\theta}_n^{\text{MLE}} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n y_i \log f(x_i, \theta) - \sum_{i=1}^n f(x_i, \theta) \quad (3)$$

2.3 Estimate calibration parameters by L_2 projection

Despite the wide applications of the least-squares approach and MLE, it can be shown that the LSE does not achieve the semiparametric efficiency when the simulator $f(x, \theta)$ is imperfect, meaning that the simulation output cannot perfectly fit the response, even with the best fit of θ . The

asymptotic variance can be reduced by the proposed estimator introduced in this section. It can also be shown that MLE is asymptotically inconsistent when the simulator $f(x, \theta)$ is imperfect. Theoretical justifications are provided in Section 3.

Assume that the number of cases y_i follows a Poisson distribution: $y_i \sim \text{Poi}(\lambda(x_i))$ for $i = 1, \dots, n$, and y_i and y_j are mutually independent for any $i \neq j$, where $\lambda(x_i)$ is the true mean function of y_i . The function $\lambda(x)$ is often called the *true process* in the computer experiment literature (Kennedy & O'Hagan, 2001; Tuo & Wu, 2015, 2016). Ideally, if the underlying mean function $\lambda(x)$ is known, the true parameter can be defined as the minimiser of the L_2 projection of the discrepancy between the true process and the simulation output, that is,

$$\theta^* = \arg \min_{\theta \in \Theta} \|\lambda(\cdot) - f(\cdot, \theta)\|_{L_2(\Omega)} \quad (4)$$

where $\|g\|_{L_2(\Omega)} = (\int_{\Omega} g(x)^2 dx)^{1/2}$.

In reality, the underlying true process $\lambda(\cdot)$ is unknown that needs to be estimated by observed data. Therefore, given the data $\{(x_i, y_i)\}_{i=1}^n$, we propose to estimate the true process by the kernel Poisson regression (Shim & Hwang, 2011; van de Geer, 2000). Similar to the conventional Poisson regression (McCullagh & Nelder, 2019), we use the logarithm as the canonical link function, that is, $\log \hat{\lambda}_n(\cdot) = \hat{\zeta}_n(\cdot)$, and $\hat{\zeta}_n(\cdot)$ is fitted by

$$\hat{\zeta}_n = \arg \min_{\zeta \in \mathcal{N}_{\Phi}(\Omega)} \frac{1}{n} \sum_{i=1}^n (\exp\{\zeta(x_i)\} - y_i \zeta(x_i)) + \kappa_n \|\zeta\|_{\mathcal{N}_{\Phi}(\Omega)}^2 \quad (5)$$

where $\|\cdot\|_{\mathcal{N}_{\Phi}(\Omega)}$ is the norm of the reproducing kernel Hilbert space generated by a given positive definite reproducing kernel Φ , and κ_n is a tuning parameter, which can be chosen by cross-validation methods. Thus, the proposed estimator of θ , which we call L_2 -estimator throughout this paper, is the minimiser of the L_2 projection as follows:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \|\hat{\lambda}_n(\cdot) - f(\cdot, \theta)\|_{L_2(\Omega)} \quad (6)$$

The optimal solution of (5) has the form of $\hat{\zeta}_n(x) = \hat{b} + \sum_{i=1}^n \hat{a}_i \Phi(x_i, x)$, where $\hat{b}$ and $\{\hat{a}_i\}_{i=1}^n$ can be obtained by the iterative re-weighted least-squares algorithm (Green & Yandell, 1985; Hastie & Tibshirani, 1990; Wahba et al., 1995). The detail of the algorithm is given in Appendix A. In practice, the calculation of the L_2 norm in (6) can be approximated by numerical integration methods, such as Monte Carlo integration (Cafisch, 1998).

As described in Section 2.1, because stochastic SEIR simulations can be quite computationally intensive, it is infeasible to obtain $f(x, \theta)$ by conducting simulations for all possible combinations of the input parameters. Thus, we employ a computationally efficient *emulator* to approximate the simulator. There are extensive studies on the development of statistical emulators in the computer experiment literature (Santner et al., 2018). GPs are the most commonly used tools in the construction of emulators (Gramacy, 2020). Based on computer experiments with sample size N , a statistical emulator is denoted by $\hat{f}_N(x, \theta)$, which produces a predictive distribution of $f(x, \theta)$ with any untried $(x, \theta) \in (\Omega, \Theta)$. Specifically, the distribution of $\hat{f}_N(x, \theta)$ with any untried $(x, \theta) \in (\Omega, \Theta)$ is a normal distribution with the mean function, defined by $m_N(x, \theta)$, and the variance function, defined by $\nu_N^2(x, \theta)$. We refer more details to Gramacy (2020). Thus, by Fubini's theorem, the L_2 -estimator of (6) can be replaced by

$$\begin{aligned} \tilde{\theta}_n &= \arg \min_{\theta \in \Theta} \mathbb{E} \|\hat{\lambda}_n(\cdot) - \hat{f}_N(\cdot, \theta)\|_{L_2(\Omega)}^2 \\ &= \arg \min_{\theta \in \Theta} \int_{\Omega} (\hat{\lambda}_n(z) - m_N(z, \theta))^2 + \nu_N^2(z, \theta) dz \end{aligned} \quad (7)$$

The applications of the proposed method with various existing emulators are demonstrated in Sections 4 and 5.

It is worth noting that the Poisson regression, $y_i \sim \text{Poi}(\lambda(x_i))$, may encounter *overdispersion* due to the presence of greater variability (McCullagh & Nelder, 2019). That is, the variance of the data is larger than the mean, which violates the assumption of Poisson distribution. The deviance goodness-of-fit test (McCullagh & Nelder, 2019) can be used to assess the model assumption. To take into account the issue of overdispersion, a *quasi-Poisson* regression can be considered which assumes that the variance of y_i is $\phi\lambda(x)$, where $\phi > 1$ is the *overdispersion parameter*. The overdispersion parameter can be estimated by the ratio of the deviance to the effective degree freedom. The details of the deviance goodness-of-fit test and the estimation of overdispersion parameter are provided in Appendix A.

3 Theoretical properties

Theoretical properties of the L_2 -estimator are discussed in this section, including the asymptotic consistency and the semiparametric efficiency. Theoretical comparisons with the LSEs are also provided by examining their asymptotic variances. The proofs are given in [online supplementary material](#).

The following theorem shows that the L_2 -estimator $\hat{\theta}_n$ in (6) is asymptotically consistent and normally distributed.

Theorem 1 Under the regularity conditions C1–C10 in Web Appendix B, we have

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}(0, 4V_0(\theta^*)^{-1}W_0(\theta^*)V_0(\theta^*)^{-1})$$

as $n \rightarrow \infty$, where

$$\begin{aligned} W_0(\theta) &= \mathbb{E} \left[\lambda(X) \frac{\partial f}{\partial \theta}(X, \theta) \frac{\partial f}{\partial \theta^T}(X, \theta) \right] \text{ and} \\ V_0(\theta) &= \mathbb{E} \left[\frac{\partial^2}{\partial \theta \partial \theta^T} (\lambda(X) - f(X, \theta))^2 \right] \end{aligned} \quad (8)$$

Remark When the overdispersion parameter, ϕ , is present in the Poisson regression, the result in Theorem 1 can be rewritten as

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}(0, 4\phi V_0(\theta^*)^{-1}W_0(\theta^*)V_0(\theta^*)^{-1})$$

By the delta method, the following corollary extends the result of Theorem 1 to a function of the L_2 -estimator, which we denote as $g(\hat{\theta}_n)$.

Corollary 2 For a function g satisfying the property that $\nabla g(\theta^*)$ exists and is nonzero valued, we have

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta^*)) \xrightarrow{d} \mathcal{N}(0, 4\nabla g(\theta^*)^T V_0(\theta^*)^{-1}W_0(\theta^*)V_0(\theta^*)^{-1}\nabla g(\theta^*))$$

as $n \rightarrow \infty$.

Corollary 2 provides a theoretical support for the estimation and inference of some commonly used quantities of interest in epidemiology, such as the basic reproduction rate, which measures the transmission potential of a disease. For instance, in the SEIR model (1), the basic reproduction rate is a ratio of two of the calibration parameters, that is, $g(\theta) = \beta/\gamma$. The result of Corollary 2 can then be applied to construct the confidence intervals for the basic reproduction rate.

When estimating the unknown parameters in compartmental models, the parameter of interest θ^* in (4) is q -dimensional, while the parameter space of the Poisson model, $y_i \sim \text{Poi}(\lambda(x_i))$, contains an infinite dimensional function space that covers λ . Therefore, the calibration problem is regarded as a semiparametric problem. For these problems, the estimation method that can reach the highest estimation efficiency is called *semiparametric efficient* (Bickel et al., 1993; Kosorok, 2008). Specifically, let Λ be an infinite dimensional parameter space whose true value is λ_0 . Suppose that T_n is an estimator for θ^* and $\sqrt{n}(T_n - \theta^*)$ is asymptotically normal. Let Λ_0 be an arbitrary finite dimensional space of Λ that satisfies $\lambda_0 \in \Lambda_0$. Consider the same calibration problem but with the parameter space Λ_0 , then under this parametric assumption and some regularity conditions, an efficient estimator can be obtained by the maximum likelihood method, which is denoted by $S_n^{\Lambda_0}$. Then, the estimator T_n is called *semiparametric efficient* if there exists a Λ_0 such that $S_n^{\Lambda_0}$ has the same asymptotic variance as T_n . More details regarding the semiparametric efficiency can be found in Tuo and Wu (2015), Bickel et al. (1993), and Kosorok (2008). It can be shown in the following theorem that the proposed L_2 -estimator is semiparametric efficient.

Theorem 3 Under the regularity conditions in Theorem 1, $\hat{\theta}_n$ is semiparametric efficient.

When the simulator f is too costly to evaluate like the stochastic SEIR simulator in Section 2.1, as discussed in Section 2.3, an statistical emulator can be considered after conducting a computer experiment of size N on the simulator. Suppose that the emulator of $f(x, \theta)$, i.e. $\hat{f}_N(x, \theta)$, follows a normal distribution with the mean function, $m_N(x, \theta)$, and the variance function, $v_N^2(x, \theta)$, i.e.

$$\hat{f}_N(x, \theta) \sim \mathcal{N}(m_N(x, \theta), v_N^2(x, \theta)) \quad (9)$$

and the L_2 -estimator is obtained by (7) as $\tilde{\theta}_n$. Then, the following theorem provides the asymptotic distribution of $\tilde{\theta}_n$.

Theorem 4 Under the regularity conditions C1 and C7–15 in Web Appendix B, we have

$$\sqrt{n}(\tilde{\theta}_n - \theta'_N) \xrightarrow{d} \mathcal{N}(0, 4V_1(\theta'_N)^{-1}W_1(\theta'_N)V_1(\theta'_N)^{-1})$$

as $n \rightarrow \infty$, where

$$\begin{aligned} \theta'_N &= \arg \min_{\theta \in \Theta} \|\lambda(\cdot) - m_N(\cdot, \theta)\|_{L_2(\Omega)}^2 + \|\sqrt{v_N^2(\cdot, \theta)}\|_{L_2(\Omega)}^2 \\ W_1(\theta) &= \mathbb{E} \left[\lambda(X) \frac{\partial m_N(X, \theta)}{\partial \theta} \frac{\partial m_N(X, \theta)}{\partial \theta^T} \right] \quad \text{and} \\ V_1(\theta) &= \mathbb{E} \left[\frac{\partial^2}{\partial \theta \partial \theta^T} \left((\lambda(X) - m_N(X, \theta))^2 + v_N^2(X, \theta) \right) \right] \end{aligned}$$

With the emulator (9), it is of no surprise that the estimator $\tilde{\theta}_n$ is asymptotic inconsistent. However, when the size of the computer experiment, N , is sufficiently large, with an appropriate emulator (e.g. GP emulator) and under some regularity conditions, we have $m_N(x, \theta) \rightarrow f(x, \theta)$ and $v_N^2(x, \theta) \rightarrow 0$ for any $(x, \theta) \in \Omega \times \Theta$ (Wang, Tuo, et al., 2020), leading to $\theta'_N \rightarrow \theta^*$, which implies that $\tilde{\theta}_n$ is asymptotic inconsistent when N is sufficiently large.

In the next theorem, the asymptotic properties of the LSE are developed and compared with those of the L_2 -estimator.

Theorem 5 Under the regularity conditions C1–C4 and C16–C17 in Web Appendix B, we have

$$\sqrt{n}(\hat{\theta}_n^{\text{LS}} - \theta^*) \xrightarrow{d} \mathcal{N}(0, 4V_0(\theta^*)^{-1}W_2(\theta^*)V_0(\theta^*)^{-1})$$

as $n \rightarrow \infty$, where

$$W_2(\theta) = W_0(\theta) + \mathbb{E} \left[(\lambda(X) - f(X, \theta))^2 \frac{\partial f}{\partial \theta}(X, \theta) \frac{\partial f}{\partial \theta^T}(X, \theta) \right]$$

Similar to the L_2 -estimator, it is shown that the LSE is asymptotically consistent and normally distributed. It can also be shown that $W_2(\theta^*) \geq W_0(\theta^*)$, which leads to

$$4V_0(\theta^*)^{-1}W_2(\theta^*)V_0(\theta^*)^{-1} \geq 4V_0(\theta^*)^{-1}W_0(\theta^*)V_0(\theta^*)^{-1} \quad (10)$$

This implies that the asymptotic variance of the LSE $\hat{\theta}_n^{\text{LS}}$ is greater or equal to that of $\hat{\theta}_n$. The equality in (10) holds if and only if

$$\mathbb{E} \left[(\lambda(X) - f(X, \theta^*))^2 \frac{\partial f}{\partial \theta}(X, \theta^*) \frac{\partial f}{\partial \theta^T}(X, \theta^*) \right] = 0 \quad (11)$$

This result indicates that if $(\partial f / \partial \theta)(x, \theta^*) \neq 0$ for all $x \in \Omega$, then (11) holds only if $\lambda(x) = f(x, \theta^*)$ for all $x \in \Omega$, which implies that the LSE $\hat{\theta}_n^{\text{LS}}$ is less efficient than $\hat{\theta}_n$ if f is an imperfect simulator, i.e. $\lambda(x) \neq f(x, \theta^*)$ for some $x \in \Omega$.

In the next theorem, the asymptotic properties of the MLE as in (3) are developed and compared with those of the L_2 -estimator.

Theorem 6 Under the regularity conditions C1 and C18–C22 in Web Appendix B, we have

$$\sqrt{n}(\hat{\theta}_n^{\text{MLE}} - \theta'') \xrightarrow{d} \mathcal{N}(0, V_3(\theta'')^{-1}W_3(\theta'')V_3(\theta'')^{-1})$$

as $n \rightarrow \infty$, where

$$\begin{aligned} \theta'' &= \arg \max_{\theta \in \Theta} \mathbb{E}[\lambda(X) \log f(X, \theta) - f(X, \theta)] \\ V_3(\theta) &= \mathbb{E} \left[\frac{1}{f(X, \theta)} \left(-\frac{1}{2} V_0(\theta) + \left(1 - \frac{\lambda(X)}{f(X, \theta)} \right) \frac{\partial f(X, \theta)}{\partial \theta} \frac{\partial f(X, \theta)}{\partial \theta^T} \right) \right] \end{aligned} \quad (12)$$

and

$$W_3(\theta) = \mathbb{E} \left[\frac{1}{f(X, \theta)^2} \left((\lambda(X) - f(X, \theta))^2 + \lambda(X) \right) \frac{\partial f(X, \theta)}{\partial \theta} \frac{\partial f(X, \theta)}{\partial \theta^T} \right]$$

Unlike the L_2 -estimator and the LSE, the MLE asymptotically converges to a value that differs from the true parameter defined in (4). For example, suppose $\lambda(x) = x^2$ and $f(x, \theta) = \theta x$, by the definition of θ^* in (4) we have $\theta^* = 0.75$, while by (12), $\hat{\theta}_n^{\text{MLE}}$ converges to $\theta'' = 2/3$ in probability.

4 Numerical study

In this section, two artificial examples are conducted to examine the finite sample performance of the proposed method and compare the estimation performance with the least-squares approach. These numerical studies are conducted on a MacBook Pro laptop with Apple M1 Max Chip and 32Gb of RAM.

4.1 Imperfect simulator with one calibration parameter

We consider an imperfect simulator adapted from Tuo and Wu (2015) with one calibration parameter. The true process is assumed to be $\lambda(x) = \exp(x/2) \sin(x/2) + 30$, where $x \in \Omega = [0, 2\pi]$, and it is illustrated in the left panel of Figure 1 as the solid line. The data are generated from equal-spaced inputs in $[0, 2\pi]$ with $n = 50$ and the outputs are generated from a Poisson distribution with the mean process $\lambda(x_i)$ for $i = 1, \dots, 50$, which are shown as the solid dots in the left panel of Figure 1.

We assume that the simulation output is $f(x, \theta) = \lambda(x) - 5\sqrt{\theta^2 - \theta + 1}(\sin(\theta x) + \cos(\theta x))$, where $\theta \in \Theta = [-1, 1]$. This simulator is imperfect because $\sqrt{\theta^2 - \theta + 1}$ is always positive for any $\theta \in \Theta$. The true parameter can be analytically solved by minimising (4), which gives that $\theta^* = -0.1789$. Plugging in the true calibration parameter, the simulator $f(x, \theta^*)$ is demonstrated as the dashed line, which is imperfect because, even with the true minimiser, the discrepancy between the simulation output and the true process is nonzero.

The performance of the L_2 -estimator is compared with the LSE and MLE based on the mean squared errors (MSEs) obtained from 100 replicates, that is, $\sum_{i=1}^{100} (\hat{\theta}_i - \theta^*)^2 / 100$, where $\hat{\theta}_i$ is the estimate at the i th replicate. Their MSEs are shown in the first three bars in the right panel of Figure 1. It shows that the L_2 -estimator ('L2') yields a smaller MSE than the 'LSE' and 'MLE'. To quantify the uncertainty of the L_2 estimator, the 95% confidence intervals are constructed based on the asymptotic result in Theorem 1, where λ , θ^* , and $\mathbb{E}$ are approximated by $\hat{\lambda}$, $\hat{\theta}_n$, and Monte-Carlo integration (Cafisch, 1998), respectively. Out of the 100 replicates, the true parameter is contained by the confidence interval 96 times, which appears to be close to the nominal coverage 95%.

We further compare the estimation performance for the cases when the simulations are computationally demanding and therefore statistical emulators are built as surrogates. Before comparing the estimation performance, we first examine the emulation performance of two existing emulation methods that are applicable to count data, which are the multiresolution functional analysis of variance emulation (Sung, Wang, et al., 2020) and the heteroscedastic GP emulation (Binois et al., 2018). Both methods have available packages in R (R Core Team, 2018), which are MRFA (Sung, 2020) and hetGP (Binois & Gramacy, 2019), respectively. These emulators are trained by conducting a computer experiment, which simulates the model outputs of $f(x, \theta)$ of size m , where the inputs are sampled from $(x, \theta) \in (\Omega, \Theta) \in \mathbb{R}^2$ using a Latin hypercube design (LHD) (McKay et al., 1979). For each input setting, simulations are conducted with a replicates. The emulation performance is examined by performing predictions on 10,000 random untried input settings from (Ω, Θ) . With four different combinations of m and a , the root mean squared prediction errors (RMSPEs) of the two emulators along with their computational time are reported in Table B1 of Appendix B. In this example, it appears that hetGP outperforms MRFA in terms of computational time and RMSPE. Thus, we select the emulator built by hetGP as the surrogate to the actual simulator in the following analysis.

Next, we compare the estimation performance with the hetGP emulator built by $m = 25$, $a = 50$ samples, leading to total sample size $N = ma = 750$. The L_2 estimator is obtained by (7) with the emulator, and the LSE is similarly obtained by minimising $\sum_{i=1}^n (y_i - m_N(x_i, \theta))^2 + \nu_N^2(z, \theta)$. For the MLE as in (3), the actual simulator $f(x, \theta)$ is replaced by the mean of the hetGP emulator, i.e. $m_N(x, \theta)$. The MSEs are shown in the last three bars in the right panel of Figure 1. Similar to the previous result without emulators, the L_2 -estimator provides a smaller MSE than the LSE and MLE. By comparing the first three and last three bars, it is not surprising to see that the MSEs of 'L2+emulator', 'LSE+emulator', and 'MLE+emulator' are larger than 'L2', 'LSE', and 'MLE' due to the prediction uncertainty from emulation. Similarly, we construct the 95% confidence intervals based on the asymptotic result in Theorem 4 for the L_2 estimator of (7), and out of the 100 replicates, the true parameter is contained by the confidence interval 91 times, which appears to be close to the nominal coverage of 95%.

4.2 Imperfect simulator with three calibration parameters

We consider a more complex problem with three calibration parameters adapted from Plumlee (2017). Assume that the true mean process is $\lambda(x) = 3x + 3x \sin(5x) + 3$ and the simulator is $f(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2$, where $x \in [0, 2]$ and $\theta \in [0, 5]^3$. Similar to the previous example, the three calibration parameters also have analytical solution $\theta^* \approx (3.56, 0.56, 1.76)$ by minimising (4).

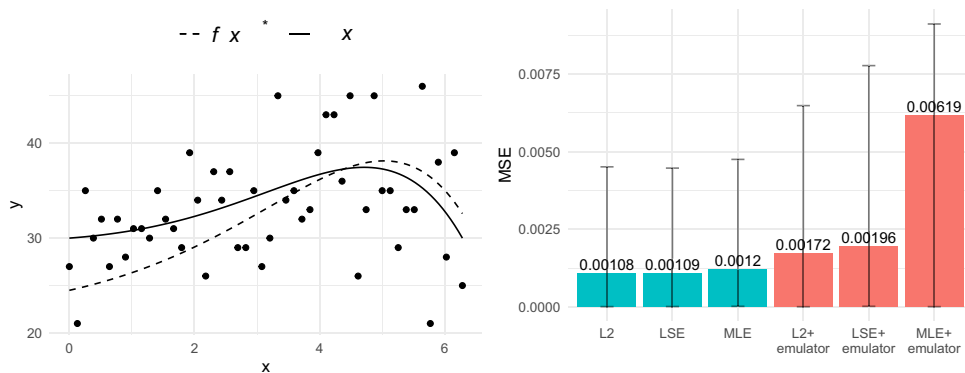


Figure 1. (Left) The true process $\lambda(x)$ as the solid line and the simulation output $f(x, \theta^*)$ as the dashed line. The real outputs are illustrated as the solid dots. (Right) MSEs of the estimates, where the error bars represent the 5% and 95% quantiles. MSE = mean squared error.

The data $\{y_i\}_{i=1}^{50}$ are generated from the Poisson distribution with the mean $\{\lambda(x_i)\}_{i=1}^{50}$, where the 50 inputs are uniformly sampled from $[0, 2]$. The estimation performance is examined based on the MSEs obtained from 100 replicates, and the proposed estimator and the LSE are compared for each calibration parameter. The results are shown in the first three bars in each plot of Figure 2, in which the y-axis represents the MSEs. Similar to the previous example, it appears that the L_2 -estimator outperforms the LSE and MLE for all of the three parameters.

In this example, we also examine the prediction performance of the two existing emulators, MRFA and hetGP. A computer experiment is conducted to train the two emulators by running the simulation outputs of $f(x, \theta)$ at m unique sample locations with a replicates, in which the unique input locations are sampled from $(x, \theta) \in (\chi, \Theta) \subseteq \mathbb{R}^4$ using an LHD. After the emulators are built, the root mean squared error are computed based on the predictions of 10,000 untried input locations, and the prediction results are summarised in Table B2 with different settings of m and a . Similar to the previous example, the hetGP method outperforms MRFA in terms of prediction accuracy and computational time. With a larger a , i.e. more replicates, the prediction accuracy of hetGP appears to increase without much increase in computational time. Thus, we select hetGP as the emulator in the following analysis.

We now compare the estimation performance for the cases where emulators are constructed as surrogates to the actual simulations. The emulator is built by hetGP with $m = 300$, $a = 100$ and based on the emulator, the estimation performance is summarised by the last three bars in each of the three plots in Figure 2. The results indicate that, either when the actual simulator is conducted or emulated, the L_2 -estimator provides smaller MSEs compared to other two estimators.

5 Analysis of COVID-19

We revisit the SEIR model in Section 2.1 and apply the proposed method to estimate the unknown parameters in the simulators for a better understanding of COVID-19 pandemic. The estimation performance based on deterministic SEIR is discussed in Section 5.1 and the stochastic version is discussed in Section 5.2. To estimate the unknown parameters, we collect the actual numbers of infected cases from Johns Hopkins University CCSE repository (Dong et al., 2020) through an R package covid19.analytics (Ponce, 2020). For each country, there are 365 observations collected from 1 March 2020 to 28 February 2021, denoted by y_i , where $i = 1, \dots, 366$. The studies are conducted for the top 20 countries which have the highest cumulative confirmed cases reported on 1 March 2021.

5.1 Parameter estimation based on deterministic SEIR

Before estimating the parameters, a deviance goodness-of-fit test is performed to examine the kernel Poisson regression as in (5), i.e. $y_i \sim \text{Poi}(\lambda_n(x_i))$. It appears that the p -values of the test are all smaller than 0.0001, which indicates that there is a lack-of-fit in the current model. Therefore, a

Table B1. Emulation performance for the example with one calibration parameter (in Section 4.1), where m is the sample size of unique locations and a is the number of replicates

Emulator	m	a	Fitting	Prediction	RMSPE
			time (s)	time (s)	
MRFA	25	50	8	0.4	9.05
	25	100	11	0.4	8.47
	50	50	11	0.7	2.31
	100	100	29	0.7	0.99
hetGP	25	50	0.15	0.02	2.08
	25	100	0.15	0.02	1.74
	50	50	0.27	0.02	1.02
	100	100	1.16	0.07	0.50

Note. Root mean squared prediction errors (RMSPEs) are reported for the two emulators based on 10,000 random predictive locations.

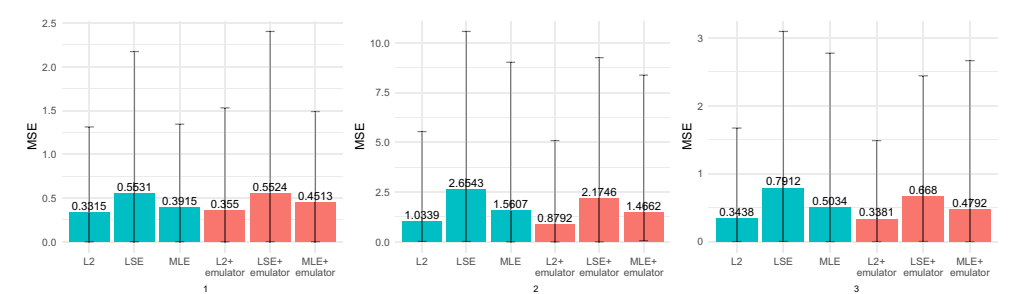


Figure 2. MSEs of the estimates of (left) θ_1 , (middle) θ_2 , and (right) θ_3 , where the error bars represent the 5% and 95% quantiles. MSE = mean squared error.

Table B2. Emulation performance for the example with three calibration parameters (in Section 4.2), where m is the sample size of unique locations and a is the number of replicates

Emulator	m	a	Fitting	Prediction	RMSPE
			time (s)	time (s)	
MRFA	300	50	258	3	0.66
	300	100	545	3	0.63
	500	5	27	2	0.82
	500	50	448	3	0.52
	500	100	29	2	0.46
hetGP	300	50	7	1	0.20
	300	100	8	1	0.16
	500	5	29	2	0.46
	500	50	29	2	0.15

Note. Root mean squared prediction errors (RMSPEs) are reported for the two emulators based on 10,000 random predictive locations.

more flexible model, the quasi-Poisson as described in Section 2.3, is applied to capture the potential overdispersion.

For each country, the L_2 -estimator of θ is obtained by minimising (6), and the corresponding estimated reproduction number R_0 can be calculated by $R_0 = \beta/\gamma$. The point estimates of R_0 and their 95% confidence intervals, which are obtained by the result of Corollary 2, are summarised in Figure 3 for the 20 countries. It shows that, from March 2020 to March 2021, all of the 20 countries have the basic reproduction numbers greater than 1, which means that the COVID-19 outbreak still poses threats to these countries. Note that the recovery rate γ is in the denominator of R_0 , and therefore the variation of R_0 appears to be higher for the countries having smaller recovery rates.

Plugging in the L_2 -estimators, the simulation results (solid lines), $f(x, \hat{\theta}_n)$, along with their confidence intervals (dashed lines), for the top 12 countries that have the highest R_0 values, are demonstrated in Figure 4. Note that the confidence intervals are similarly constructed based on Corollary 2. That is, the variance of $f(x, \hat{\theta}_n)$ can be approximated by

$$4\nabla_{\theta}f(x, \hat{\theta}_n)^T V_0(\hat{\theta}_n)^{-1} W_0(\hat{\theta}_n) V_0(\hat{\theta}_n)^{-1} \nabla_{\theta}f(x, \hat{\theta}_n) \quad (13)$$

where ∇_{θ} is the partial derivative with respect to θ . In general, it appears that the simulation results can reasonably capture the overall trend observed from the actual numbers of infected cases, which are shown as dots. For Iran, Czechia, and Spain, the discrepancy between the simulation results and actual observations is relatively larger than the other countries. This is partly because SEIR is an imperfect simulator which is built based on some assumptions or simplifications, and these assumptions may have larger deviations from the reality for certain countries. Another reason is that the intrinsic dynamics are neglected in the deterministic simulations. To take into account the dynamics, a stochastic simulator is considered in the next section.

5.2 Parameter estimation based on stochastic SEIR

Conducting stochastic simulations based on SEIR is computationally intensive, therefore emulators are developed as a faster surrogate to the actual stochastic simulations. In this study, we consider the `hetGP` emulator, which is built based on the simulations generated using a 60-run LHD for parameter settings with 20 equal-spaced time steps in x , which leads to the total sample size of $m = 1,200$. For each parameter-input setting, 50 replicates are simulated, i.e. $a = 50$, so the total sample size of this computer experiment is $N = ma = 60,000$. Based on this emulator, it takes less than 2 s to emulate the result for an untried parameter setting, which is significantly faster than the actual stochastic simulation.

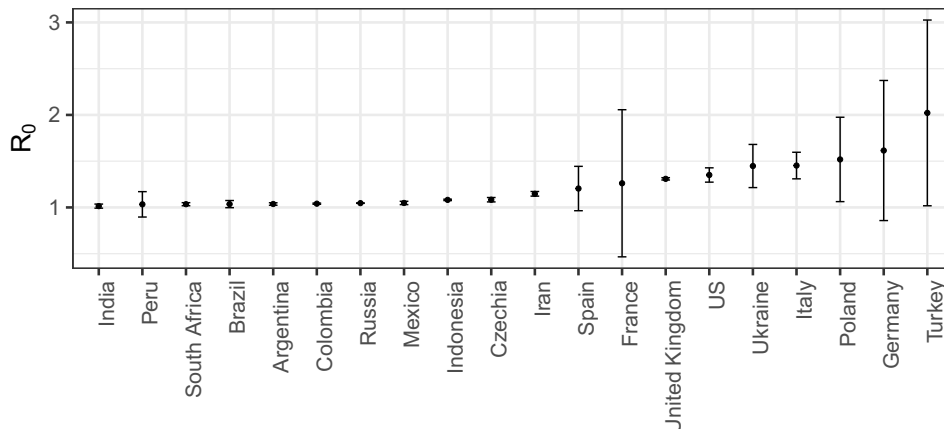


Figure 3. The estimated reproduction numbers for top 20 infectious countries based on the deterministic SEIR model. SEIR = susceptible-exposed-infectious-recovered.

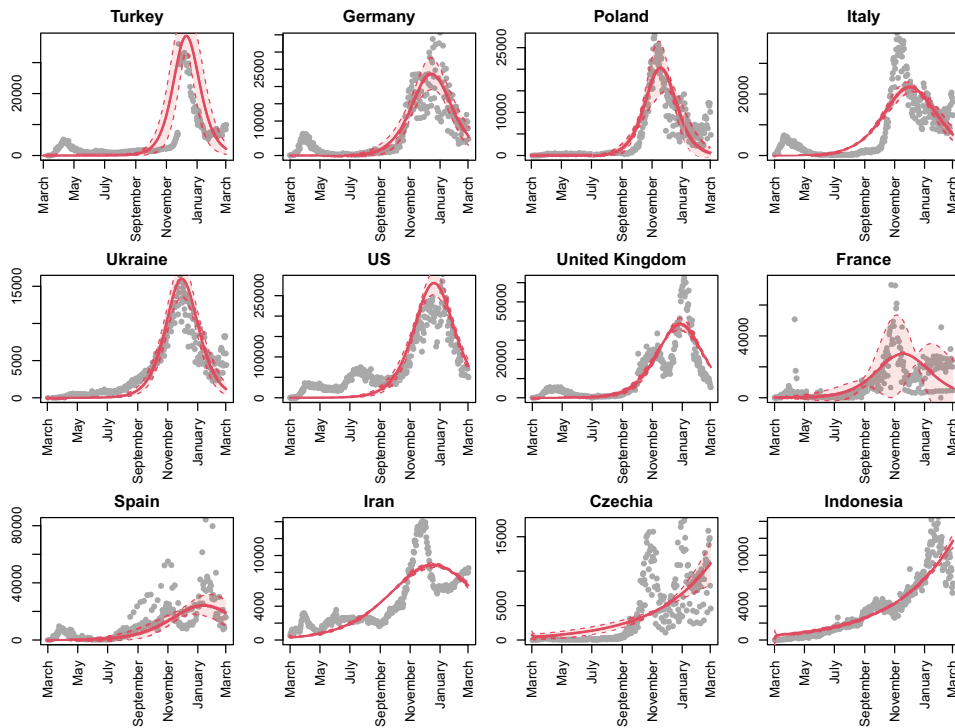


Figure 4. The solid circle dots are the actual numbers of daily infected cases. The solid lines are the results from deterministic SEIR simulators by plugging in the L_2 estimates, and the dashed lines are their corresponding 95% confidence intervals. SEIR = susceptible-exposed-infectious-recovered.

With the *hetGP* emulator, which has the form of (9), the L_2 -estimators are obtained by minimising (7). The corresponding estimates of R_0 and their 95% confidence intervals are summarised in Figure 5, where the variance is obtained based on the result of Theorem 4. It appears that South Africa and Argentina have their basic reproduction numbers controlled below 0.9, which also show small basic reproduction numbers in the deterministic simulations (less than 1.05). We further report the estimated incubation period, $1/\kappa$, for each country and the corresponding 95% confidence intervals in Figure 6. The overall average incubation period is 5.15 as indicated by the dashed line. When comparing with the deterministic version, the estimation uncertainty based on the stochastic model is smaller. For example, the confidence intervals in Figure 5 are generally narrower than the ones in Figure 3. The main reason is that the stochastic SEIR model accounts for the randomness and therefore the estimation is more robust to the noise, which leads to smaller uncertainty in the R_0 values compared to its deterministic counterpart. Having a slightly larger sample size for some countries may also be a factor of smaller uncertainty. Furthermore, we employed a frequentist framework and plugged the point estimate in the asymptotic variance in Corollary 2, which may lead to an underestimation of the uncertainty from parameter estimation. To address this concern, an alternative approach is to adopt a Bayesian framework that incorporates prior distributions on the parameters. Further discussions regarding this Bayesian framework can be found in Section 6.

In Figure 7, the actual numbers of infected cases are illustrated as dots. By plugging in the L_2 -estimators, the simulation results for the top 12 countries with the highest R_0 are illustrated as curves, along with the 95% confidence intervals as dashed lines. Overall, the simulation results show a much better agreement with the actual observations compared to the deterministic ones in Section 5.1. In particular, by taking into account the intrinsic dynamics, the simulation discrepancy for Czechia is significantly reduced from the deterministic one shown in Figure 4. Note that the

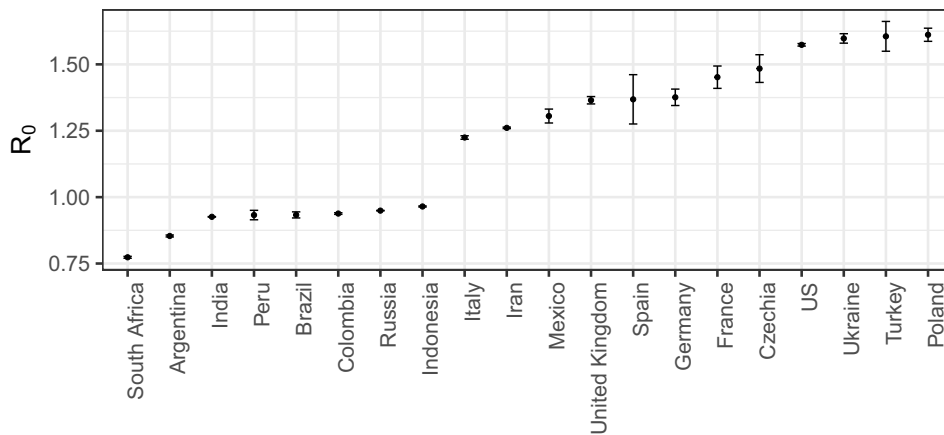


Figure 5. The reproduction numbers of top 20 infectious countries based on the stochastic SEIR model. SEIR = susceptible-exposed-infectious-recovered.

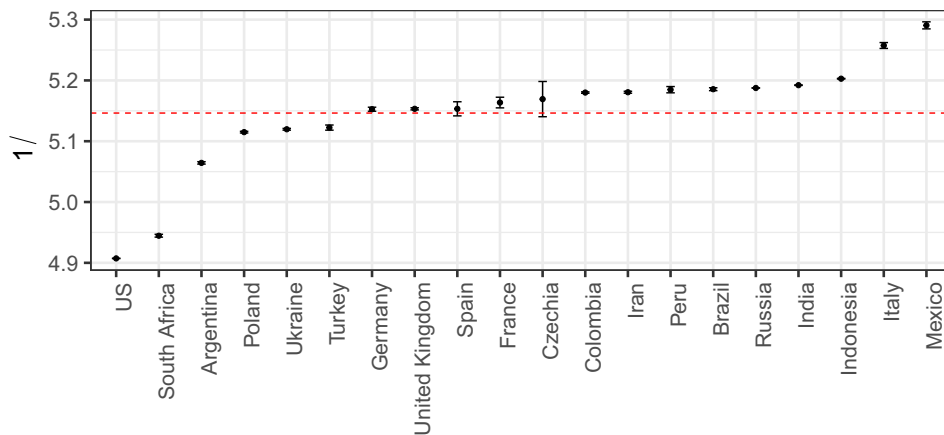


Figure 6. The estimated average incubation period based on the stochastic SEIR model, along with the overall average as indicated by the dashed line. SEIR = susceptible-exposed-infectious-recovered.

confidence intervals are computed based on $\mathbb{V}[\hat{f}_N(x, \tilde{\theta})] = \mathbb{E}[\mathbb{V}[\hat{f}_N(x, \tilde{\theta}_n) | \tilde{\theta}_n]] + \mathbb{V}[\mathbb{E}[\hat{f}_N(x, \tilde{\theta}_n) | \tilde{\theta}_n]]$, which can be approximated by

$$\nu_N^2(x, \tilde{\theta}_n) + 4\nabla_{\theta} m_N(x, \tilde{\theta}_n)^T V_1(\tilde{\theta}_n)^{-1} W_1(\tilde{\theta}_n) V_1(\tilde{\theta}_n)^{-1} \nabla_{\theta} m_N(x, \tilde{\theta}_n) \quad (14)$$

using the result of Theorem 4. When comparing with the predictive uncertainty of the deterministic model as shown in (13), the stochastic version as in (14) introduces an additional source of uncertainty captured by the term $\nu_N^2(x, \tilde{\theta}_n)$, which accounts for the uncertainty due to emulation. This term contributes a dominating effect to the overall uncertainty, especially when stochastic models are computationally expensive and the emulators are constructed based on a limited number of computer experiments. As a result, even though the estimation uncertainty is relatively smaller with the stochastic model, the predictive uncertainty presented in Figure 7 is generally wider than the ones from the deterministic SEIR in Figure 4.

6 Discussions and concluding remarks

Epidemic models for the analysis of COVID-19 are often imperfect. A new calibration method is proposed to estimate the unknown parameters in the imperfect epidemic models. The proposed

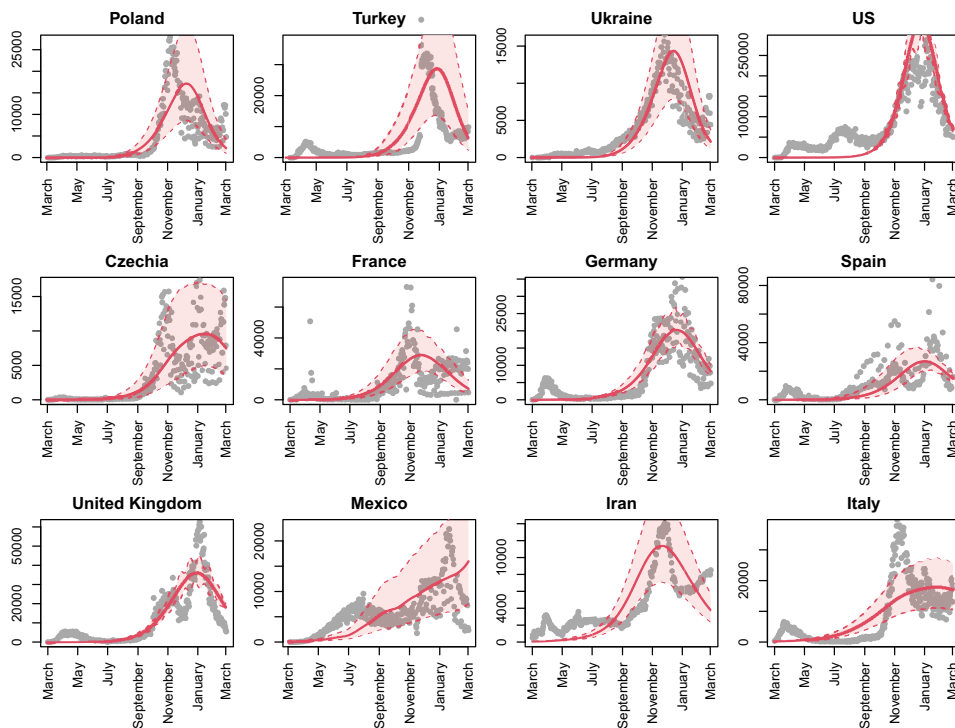


Figure 7. Number of infectious (solid circle dots) and the best fit of the stochastic SEIR models (solid lines) of top 12 most infectious countries, where the dashed lines are their corresponding 95% confidence intervals. SEIR = susceptible-exposed-infectious-recovered.

estimator outperforms the LSE by providing a smaller estimation variance and achieving the semi-parametric efficiency. The proposed method is applied to the SEIR model for the analysis of COVID-19 pandemic. The estimates of the quantities of interest, such as the basic reproduction number and the average incubation period, and their confidence intervals are obtained based on the asymptotic results.

Apart from the frequentist approach studied in this paper, we are currently developing a Bayesian framework that extends the recent developments of Bayesian calibration to count data. For example, the orthogonal GP models (Plumlee et al., 2016) or the Bayesian projected calibration (Tuo, 2019; Xie & Xu, 2021) can be used to model the model discrepancy, which addresses the unidentifiability issue for continuous outputs, and it is conceivable to further extend the modelling to count data by incorporating the idea of the generalised calibration in Grosskopf et al. (2020). This framework is particularly useful when the goal is to provide a better fit to the data. Moreover, by incorporating prior distributions on the parameters and allowing for a range of plausible values, a Bayesian analysis can provide a more comprehensive assessment of uncertainty of the estimates. It is also worth investigating the confidence set on the calibration parameters using the method of Plumlee (2019) for the application herein. Another interesting direction that deserves further studies is to relax the constant parameter assumption. Instead, the calibration parameters can be assumed to be functions of some factors, such as time or temperature, which not only increases the model flexibility but also can provide further insights to the time-course dynamics of the COVID-19 infection.

Acknowledgments

The authors gratefully acknowledge helpful advice from the associate editor and two anonymous referees whose comments greatly strengthen this paper.

Conflict of interest: No conflicts of interest to declare.

Funding

This work was supported by NSF DMS 1660477 and NSF HDR TRIPODS award CCF 1934924.

Data availability

The code and data can be accessed on the GitHub repository located at the following link: <https://github.com/ChihLi/Epidemic-Models-Calibration>

Supplementary material

[Supplementary material](#) is available online at *Journal of the Royal Statistical Society: Series C*.

Appendix A. Algorithm to Estimate ξ in (3) and Estimate Overdispersion Parameter ϕ

Since the optimal solution has the form of $\xi_n(x) = b + \sum_{i=1}^n a_i \Phi(x_i, x)$, one can show that the penalised likelihood in (5) can be rewritten as

$$\frac{1}{n} \sum_{i=1}^n \{ \exp(b + \mathbf{a}^T \boldsymbol{\psi}(x_i)) - y_i(b + \mathbf{a}^T \boldsymbol{\psi}(x_i)) \} + \kappa_n \mathbf{a}^T \boldsymbol{\Phi} \mathbf{a}$$

where $\mathbf{a} = (a_1, \dots, a_n)$, $\boldsymbol{\psi}(x) = (\Phi(x, x_1), \dots, \Phi(x, x_n))$, and $\boldsymbol{\Phi} = (\Phi(x_i, x_j))_{1 \leq i, j \leq n}$. The optimal solution of $\mathbf{a}$ and b can then be obtained by taking the first-order partial derivatives of the objective function with respect to $\mathbf{a}$ and b and setting them equal to zero, which can be solved by the iterative re-weighted least-squares algorithm as follows. Denote

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 0 & 0_n^T \\ 0_n & \boldsymbol{\Phi} \end{pmatrix}, \quad \boldsymbol{\Phi}_1 = \begin{pmatrix} \mathbf{1}_n & \boldsymbol{\Phi} \end{pmatrix}$$

where $\mathbf{1}_n = [1, \dots, 1]^T$ and $0_n = [0, \dots, 0]^T$, and denote $\mathbf{W}$ as an $n \times n$ diagonal matrix with diagonal elements $\mathbf{W}_{ii} = \exp(b + \mathbf{a}^T \boldsymbol{\psi}(x_i))$. Then, in each step, one first solves for $\boldsymbol{\beta} := (b, \mathbf{a}^T)^T$ in

$$(\boldsymbol{\Phi}_1^T \mathbf{W} \boldsymbol{\Phi}_1 + 2n\kappa_n \boldsymbol{\Phi}_0) \boldsymbol{\beta} = \boldsymbol{\Phi}_1^T \mathbf{W} \boldsymbol{\eta}$$

with an initial guess of $\boldsymbol{\eta}$, which is a vector of size n , and then update each element of $\boldsymbol{\eta}$ by

$$\eta_i = (b + \mathbf{a}^T \boldsymbol{\psi}(x_i)) + \frac{y_i - \exp(b + \mathbf{a}^T \boldsymbol{\psi}(x_i))}{\exp(b + \mathbf{a}^T \boldsymbol{\psi}(x_i))}$$

The estimate $\hat{\boldsymbol{\beta}}$ can then be obtained by continuing solving for $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ iteratively until some convergence criterion is met.

To examine the goodness-of-fit of the Poisson regression, the following deviance goodness-of-fit test is considered. Since it can be shown that the deviance of the model follows a chi-square distribution asymptotically, that is

$$D = 2 \sum_{i=1}^n (y_i \log(y_i / \hat{\lambda}_n(x_i)) - (y_i - \hat{\lambda}_n(x_i))) \xrightarrow{d} \chi_{\text{edf}}$$

when n is sufficiently large, where the effective degree freedom, $\text{edf} = \text{trace}(\mathbf{S})$, where

$$\mathbf{S} = \boldsymbol{\Phi}_1 (\boldsymbol{\Phi}_1^T \mathbf{W} \boldsymbol{\Phi}_1 + 2n\kappa_n \boldsymbol{\Phi}_0)^{-1} \boldsymbol{\Phi}_1^T \mathbf{W}$$

If the test indicates that overdispersion is present in the Poisson model, the overdispersion parameter ϕ can be estimated by $\hat{\phi} = D/\text{edf}$.

Appendix B. Numerical Comparison of Emulators

The numerical comparisons of the two emulators, MRFA and `hetGP`, for the numerical studies in Sections 4.1 and 4.2 are given in this section.

References

- Allen L. J. S. (2008). An introduction to stochastic epidemic models. In F. Brauer, P. van den Driessche, & J. Wu (Eds.), *Mathematical epidemiology* (pp. 81–130). Springer.
- Allen L. J. S. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2), 128–142. <https://doi.org/10.1016/j.idm.2017.03.001>
- Anastassopoulou C., Russo L., Tsakris A., & Siettos C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One*, 15(3), e0230405. <https://doi.org/10.1371/journal.pone.0230405>
- Andersson H., & Britton T. (2012). *Stochastic epidemic models and their statistical analysis*. Springer Science & Business Media.
- Annas S., Pratama M. I., Rifandi M., Sanusi W., & Side S. (2020). Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. *Chaos, Solitons & Fractals*, 139, 110072. <https://doi.org/10.1016/j.chaos.2020.110072>
- Bayarri M. J., Berger J. O., Paulo R., Sacks J., Cafeo J. A., Cavendish J., Lin C.-H., & Tu J. (2007). A framework for validation of computer models. *Technometrics*, 49(2), 138–154. <https://doi.org/10.1198/004017007000000092>
- Bentout S., Chekroun A., & Kuniya T. (2020). Parameter estimation and prediction for coronavirus disease outbreak 2019 (COVID-19) in Algeria. *AIMS Public Health*, 7(2), 306–318. <https://doi.org/10.3934/publichealth.2020026>
- Bickel P. J., Klaassen C. A. J., Ritov Y., & Wellner J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press.
- Binois M., & Gramacy R. B. (2019). *hetGP: Heteroskedastic Gaussian process modeling and design under replication* (r package version 1.1.1). <https://CRAN.R-project.org/package=hetGP>
- Binois M., Gramacy R. B., & Ludkovski M. (2018). Practical heteroscedastic Gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical Statistics*, 27(4), 808–821. <https://doi.org/10.1080/10618600.2018.1458625>
- Caffisch R. E. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7(1), 1–49. <https://doi.org/10.1017/S0962492900002804>
- Capaldi A., Behrend S., Berman B., Simth J., Wright J., & Lloyd A. L. (2012). Parameter estimation and uncertainty quantification for an epidemic model. *Mathematical Biosciences and Engineering*, 9(3), 553–576. <https://doi.org/10.3934/mbe.2012.9.553>
- Carcione J. M., Santos J. E., Bagaini C., & Ba J. (2020). A simulation of a COVID-19 epidemic based on a deterministic SEIR model. *Frontiers in Public Health*, (8). <https://doi.org/10.3389/fpubh.2020.00230>
- Chen X., & Qiu Z. (2020). Scenario analysis of non-pharmaceutical interventions on global COVID-19 transmissions. *Covid Economics: Vetted and Real-Time Papers*, (7), 46–67. <https://cepr.org/node/390388>
- Chowell G. (2017). Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3), 379–398. <https://doi.org/10.1016/j.idm.2017.08.001>
- Chowell G., Castillo-Chavez C., Fenimore P. W., Kribs-Zaleta C. M., Arriola L., & Hyman J. M. (2004). Model parameters and outbreak control for SARS. *Emerging Infectious Diseases*, 10(7), 1258–1263. <https://doi.org/10.3201/eid1007.030647>
- Chowell G., Fenimore P. W., Castillo-Garsow M. A., & Castillo-Chavez C. (2003). SARS outbreaks in Ontario, Hong Kong and Singapore: The role of diagnosis and isolation as a control mechanism. *Journal of Theoretical Biology*, 224(1), 1–8. [https://doi.org/10.1016/S0022-5193\(03\)00228-5](https://doi.org/10.1016/S0022-5193(03)00228-5)
- Diekmann O., Heesterbeek J. A. P., & Britton T. (2013). *Mathematical tools for understanding infectious disease dynamics*. Princeton University Press.
- Dong E., Du H., & Gardner L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5), 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Epstein J. M. (2009). Modelling to contain pandemics. *Nature*, 460(7256), 687–687. <https://doi.org/10.1038/460687a>
- Farah M., Birrell P., Conti S., & Angelis D. D. (2014). Bayesian emulation and calibration of a dynamic epidemic model for A/H1N1 influenza. *Journal of the American Statistical Association*, 109(508), 1398–1411. <https://doi.org/10.1080/01621459.2014.934453>

- Funk S., Gilad E., Watkins C., & Jansen V. A. A. (2009). The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16), 6872–6877. <https://doi.org/10.1073/pnas.0810762106>
- Gillespie D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25), 2340–2361. <https://doi.org/10.1021/j100540a008>
- Giordano G., Blanchini F., Bruno R., Colaneri DFA P., Di Matteo A., & Colaneri M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 26(6), 855–860. <https://doi.org/10.1038/s41591-020-0883-7>
- Gramacy R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC Press.
- Gramacy R. B., Bingham D., Holloway J. P., Grosskopf M. J., Kuran C. C., Rutter E., Trantham M., & Drake R. P. (2015). Calibrating a large computer experiment simulating radiative shock hydrodynamics. *The Annals of Applied Statistics*, 9(3), 1141–1168. <https://doi.org/10.1214/15-AOAS850>
- Green P. J., & Yandell B. S. (1985). Semi-parametric generalized linear models. In *Proceedings 2nd international glim conference* (pp. 44–55). Lancaster, Lecture notes in statistics No. 32. Springer.
- Grosskopf M., Bingham D., Adams M. L., Hawkins W. D., & Perez-Nunez D. (2021). Generalized computer model calibration for radiation transport simulation. *Technometrics*, 63(1), 27–39. <https://doi.org/10.1080/00401706.2019.1701557>
- Han G., Santner T. J., & Rawlinson J. J. (2009). Simultaneous determination of tuning and calibration parameters for computer experiments. *Technometrics*, 51(4), 464–474. <https://doi.org/10.1198/TECH.2009.08126>
- Hastie T., & Tibshirani R. (1990). *Generalized additive models*. Chapman and Hall.
- He S., Peng Y., & Sun K. (2020). SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics*, 101, 1667–1680. <https://doi.org/10.1007/s11071-020-05743-y>
- Heesterbeek H., Anderson R. M., Andreasen V., Bansal S., De Angelis D., Dye C., Eames K. T. D., Edmunds W. J., Frost S. D. W., Funk S., Hollingsworth T. D., House T., Isham V., Klepac P., Lessler J., Lloyd-Smith J. O., Metcalf C. J. E., Mollison D., Pellis L., ... , Isaac Newton Institute IDD Collaboration (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227), aaa4339. <https://doi.org/10.1126/science.aaa4339>
- Hindmarsh A. C. (1980). LSODE and LSODI, two new initial value ordinary differential equation solvers. *ACM Sigum Newsletter*, 15(4), 10–11. <https://doi.org/10.1145/1218052.1218054>
- Hodges J. S., & Riech B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4), 325–334. <https://doi.org/10.1198/tast.2010.10052>
- Kennedy M. C., & O'Hagan A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B*, 63(3), 425–464. <https://doi.org/10.1111/1467-9868.00294>
- Kosorok M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.
- McCullagh P., & Nelder J. A. (2019). *Generalized linear models* (2nd ed.). Routledge.
- McKay M. D., Beckman R. J., & Conover W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239–245. <https://doi.org/10.2307/1268522>
- Mwalili S., Kimanthi M., Ojiambo V., Gathungu D., & Mbogo R. W. (2020). SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Research Notes*, 13(1), 352. <https://doi.org/10.1186/s13104-020-05192-1>
- Paciorek C. J. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, 25(1), 107–125. <https://doi.org/10.1214/10-STS326>
- Plumlee M. (2017). Bayesian calibration of inexact computer models. *Journal of the American Statistical Association*, 112(519), 1274–1285. <https://doi.org/10.1080/01621459.2016.1211016>
- Plumlee M. (2019). Computer model calibration with confidence and consistency. *Journal of the Royal Statistical Society: Series B*, 81(3), 519–545. <https://doi.org/10.1111/rssb.12314>
- Plumlee M., Joseph V. R., & Yang H. (2016). Calibrating functional parameters in the ion channel models of cardiac cells. *Journal of the American Statistical Association*, 111(514), 500–509. <https://doi.org/10.1080/01621459.2015.1119695>
- Ponce M. (2020). *Covid19.analytics: Load and analyze live data from the CoViD-19 pandemic* (r package version 1.1). <https://CRAN.R-project.org/package=covid19.analytics>
- R Core Team (2018). *R: A language and environment for statistical computing*. R foundation for statistical computing. <https://www.R-project.org/>
- Santner T. J., Williams B. J., & Notz W. I. (2018). *The design and analysis of computer experiments* (2nd ed.). Springer.
- Shim J., & Hwang C. (2011). Kernel Poisson regression machine for stochastic claims reserving. *Journal of the Korean Statistical Society*, 40(1), 1–9. <https://doi.org/10.1016/j.jkss.2010.01.004>

- Sung C. L. (2020). *MRFA: Fitting and predicting large-scale nonlinear regression problems using multi-resolution functional ANOVA (MRFA) approach* (r package version 0.4). R Package. <https://CRAN.R-project.org/package=MRFA>
- Sung C. L., Hung Y., Rittase W., Zhu C., & Wu C. F. J. (2020). A generalized Gaussian process model for computer experiments with binary time series. *Journal of the American Statistical Association*, 115(530), 945–956. <https://doi.org/10.1080/01621459.2019.1604361>
- Sung C. L., Wang W., Plumlee M., & Haaland B. (2020). Multiresolution functional ANOVA for large-scale, many-input computer experiments. *Journal of the American Statistical Association*, 115(530), 908–919. <https://doi.org/10.1080/01621459.2019.1595630>
- Tuo R. (2019). Adjustments to computer models via projected kernel calibration. *SIAM/ASA Journal on Uncertainty Quantification*, 7(2), 553–578. <https://doi.org/10.1137/17M1128769>
- Tuo R., & Wu C. F. J. (2015). Efficient calibration for imperfect computer models. *The Annals of Statistics*, 43(6), 2331–2352. <https://doi.org/10.1214/15-AOS1314>
- Tuo R., & Wu C. F. J. (2016). A theoretical framework for calibration in computer models: Parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1), 767–795. <https://doi.org/10.1137/151005841>
- van de Geer S. (2000). *Empirical processes in M-estimation*. Cambridge University Press.
- Wahba G., Gu C., Wang Y., & Campbell R. (1995). Soft classification, a.k.a. risk estimation, via penalized log likelihood and smoothing spline analysis of variance. In D. H. Wolpert (Ed.), *The mathematics of generalization* (pp. 329–360). Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley.
- Wang L., Zhou Y., He J., Wang F., Tang E. M. L., & Song P. (2020). *An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China*, MedRxiv preprint.
- Wang W., Tuo R., & Jeff Wu C. F. (2020). On prediction properties of kriging: Uniform error bounds and robustness. *Journal of the American Statistical Association*, 115(530), 920–930. <https://doi.org/10.1080/01621459.2019.1598868>
- Widgren S., Bauer P., Eriksson R., & Engblom S. (2019). SimInf: An R package for data-driven stochastic disease spread simulations. *Journal of Statistical Software*, 91(12), 1–42. <https://doi.org/10.18637/jss.v091.i12>
- Wu J. T., Leung K., & Leung G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet*, 395(10225), 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
- Xie F., & Xu Y. (2021). Bayesian projected calibration of computer models. *Journal of the American Statistical Association*, 116(536), 1965–1982. <https://doi.org/10.1080/01621459.2020.1753519>