*Submitted to the Annals of Applied Statistics*

# ESTIMATING FUNCTIONAL PARAMETERS FOR UNDERSTANDING THE IMPACT OF WEATHER AND GOVERNMENT INTERVENTIONS ON COVID-19 OUTBREAK

BY CHIH-LI SUNG

*Department of Statistics and Probability, Michigan State University, sungchih@msu.edu*

As the coronavirus disease 2019 (COVID-19) has shown profound effects on public health and the economy worldwide, it becomes crucial to assess the impact on the virus transmission and develop effective strategies to address the challenge. A new statistical model derived from the SIR epidemic model with functional parameters is proposed to understand the impact of weather and government interventions on the virus spread in the presence of asymptomatic infections among eight metropolitan areas in the United States. The model uses Bayesian inference with Gaussian process priors to study the functional parameters nonparametrically, and sensitivity analysis is adopted to investigate the main and interaction effects of these factors. This analysis reveals several important results including the potential interaction effects between weather and government interventions, which shed new light on the effective strategies for policymakers to mitigate the COVID-19 outbreak.

**1. Introduction.** As the coronavirus disease 2019 (COVID-19) has already had profound effects on public health and the economy worldwide, how to use statistical approaches to model and understand the spread of COVID-19 to inform and educate the public about the virus transmission and develop effective strategies for addressing this challenge has become crucial. In particular, the understanding of how government interventions and environmental factors, such as temperature and humidity, affect the virus transmissibility is important yet unclear. Moreover, an effective strategy to mitigate the outbreak based on the weather conditions is in extreme need for policymakers yet little attention has been paid to the interaction effect between weather and government interventions. For instance, a natural question for policymakers is "Should the government implement more restrictions to mitigate the pandemic as the weather gets colder?"

Since the COVID-19 outbreak, many studies have investigated the impact of weather and government interventions, but some challenges remain. Yu (2020); Xu et al. (2020); Carson et al. (2020) find some evidence that the weather may be associated with the COVID-19 spread, while Jamil et al. (2020); Gupta, Pradhan and Maulud (2020) find no significant associations. Most of the studies on the impacts of government interventions on COVID-19 spread show that government interventions are associated with reduced COVID-19 transmission; e.g., Cowling et al. (2020); Haug et al. (2020); Haldar and Sethi (2020); Flaxman et al. (2020). However, most of this work focuses on individual effects of weather and government interventions, which may lead to misleading results due to potential collinearity issues (Wilson, 2020). For example, cold weather may increase the risk of disease overall, leading governments to impose travel restrictions. Further, interaction effects between weather and government interventions variables cannot be estimated if the effects in these two sets of variables are estimated separately. Moreover, most of the existing work study the impact *without* accounting for the presence of the persons who are asymptomatic but can nevertheless infect others, which is a distinguishing feature of COVID-19 in comparison with previous viral diseases.

*Keywords and phrases:* Basic reproduction number, Asymptomatic infections, Epidemic model, Nonparametric regression, Sensitivity analysis.

In this paper, we employ a nonparametric regression method not only to model the impacts of weather and government interventions *jointly* in the presence of asymptomatic infections, which incorporates an epidemic model that allows us to evaluate the effects on the virus transmissibility, but also provide forecasts of future COVID-19 infections. Specifically, a Gaussian process prior (Williams and Rasmussen, 2006) is imposed on the functional parameters in the *susceptible-infectious-removed (SIR)* model (Kermack and McKendrick, 1927), and based on this model, the posterior distribution of the basic reproduction number, which is used to measure the transmission potential of a disease, will be derived. Both main and interaction effects of these factors will be analyzed by sensitivity analysis (Sobol, 1993).

Parameter estimation in epidemic models is often called *calibration* in the computer experiment literature (Kennedy and O'Hagan, 2001; Santner, Williams and Notz, 2018; Tuo and Wu, 2015). Although there are numerous developments on calibration, most of the existing work are based on scalar parameters rather than functional parameters. Exceptions include the recent work by Plumlee, Joseph and Yang (2016); Brown and Atamturktur (2018), but their work is based on continuous outputs with a Gaussian assumption, which does not hold for count data in the epidemic models.

In Section 2, the SIR model and a modified SIR model with functional parameters will be introduced. The statistical model incorporated with the SIR model will be explicitly described in Section 3. Numerical studies are conducted in Section 4 to examine the performance. In Section 5, the statistical model is applied to the COVID-19 outbreak to assess the impacts of weather and government interventions. Final remarks are given in Section 6. The details of sampling for the posterior distributions, the R (R Core Team, 2018) code, and the data for reproducing the results in this paper are provided in the Supplementary Material (Sung, 2022).

## 2. Compartmental models in epidemiology.

2.1. *SIR model.* Compartmental models are widely used in epidemiology which simplify the mathematical modelling of infectious diseases. One of the prominent models is the *SIR model* (Kermack and McKendrick, 1927; Diekmann, Heesterbeek and Britton, 2012), which assigns the population to three compartments: susceptible (S), infectious (I), and removed (R), where the three compartments respectively represent the number of the susceptible individuals, the infected individuals, and the removed individuals, which include the ones who are recovered, quarantined or deceased. The SIR model has been widely used for understanding how a disease spreads in outbreaks of measles, influenza, rubella, smallpox, Ebola, monkeypox, SARS, and the current COVID-19 pandemic. See, for example, Osthus et al. (2017); Chen et al. (2020); Cooper, Mondal and Antonopoulos (2020); Roda et al. (2020); D'Arienzo and Coniglio (2020).

Transitions among the three compartments can be expressed mathematically by three ordinary differential equations as follows:

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)S(t)}{N}, \quad \frac{dI(t)}{dt} = \frac{\beta I(t)S(t)}{N} - \gamma I(t), \quad \frac{dR(t)}{dt} = \gamma I(t), \tag{1}$$

where $S(t)$, $I(t)$ and $R(t)$ represent the numbers of cases in the corresponding compartments, $N = S(t) + I(t) + R(t)$ is the total population, $\beta$ is the contact rate that represents the average number of contacts per person per time in the susceptible compartment that is sufficient to spread the disease, and $\gamma$ is the removed rate from the infectious compartment to the removed compartment. The ratio of $\beta$ and $\gamma$ is called the *basic reproduction number* in epidemiology, often denoted by $\mathcal{R}_0 := \beta/\gamma$, which indicates the average number of infected cases generated by a typical infectious individual when introduced into a fully susceptible

population. This number is of great importance in public health and epidemiology, which is often used to measure the transmission potential of a disease or a virus (Dietz, 1993; Zhao et al., 2020; Zhang et al., 2020). Essentially, when $\mathcal{R}_0$ is larger than 1, the infection will be able to start spreading in a population, and the larger $\mathcal{R}_0$ is, the harder it is to control the epidemic.

2.2. *Modified SIR model.* Although the SIR model has been widely used in epidemiology, the model has been shown that it cannot reflect the reality due to its simplifications and assumptions. See, for example, Heesterbeek et al. (2015); Sung and Hung (2020). In particular, the constant parameter assumption of $\beta$ and $\gamma$, which implies that the contact rate and the removed rate are both fixed in the entire process, is too strong and unrealistic (Cowling, Ho and Leung, 2008; Cauchemez et al., 2016; Hong and Li, 2020; Yu, 2020; Ambrosio and Aziz-Alaoui, 2020). Therefore, in this article, we consider a modified, more flexible SIR model by assuming that the parameters can vary based on potential factors.

First, similar to Hong and Li (2020), we consider a discrete version of SIR models by replacing the derivatives in (1) with finite differences, which results in

$$I(t+1) - I(t) = \frac{\beta I(t)(N - I(t) - R(t))}{N} - \gamma I(t),$$
$$R(t+1) - R(t) = \gamma I(t).$$

Then, by assuming the functional parameters $\beta(\mathbf{x})$ and $\gamma(\mathbf{x})$, where $\mathbf{x} \in \Omega \subseteq \mathbb{R}^d$ is a $d$-dimensional factor, and expressing the equations in a recursive fashion, a modified SIR model can be expressed as

$$I(t+1) = (1 + \beta(\mathbf{x}) - \gamma(\mathbf{x}))I(t) + \beta(\mathbf{x})I(t)(I(t) + R(t))/N,$$
$$R(t+1) = R(t) + \gamma(\mathbf{x})I(t),$$

and $S(t+1) = N - I(t+1) - R(t+1)$ for $t \in \mathbb{N} \cup \{0\}$. Thus, the number of the daily infectious cases at day $t$ based on the modified SIR is the difference in susceptible from day $t-1$ to day $t$, which we denoted as

$$f(t, \beta(\mathbf{x}), \gamma(\mathbf{x})) := S(t-1) - S(t). \tag{2}$$

## 3. A statistical model incorporated with the SIR model.

3.1. *Gaussian process priors for functional parameters.* In this section, we introduce a statistical model incorporated with the modified SIR model in (2). First, denote $y_t$ as the daily reported number of infectious cases at day $t$, and assume $y_t$ follows an independent Poisson distribution with the mean function that is a fraction of $f(t, \beta(\mathbf{x}), \gamma(\mathbf{x}))$. That is,

$$y_t \overset{\text{indep.}}{\sim} \text{Poi}(\kappa(t) f(t, \beta(\mathbf{x}), \gamma(\mathbf{x}))), \tag{3}$$

where $\kappa(t) \in (0, 1)$, indicating that only a fraction $\kappa(t)$ of the total number of infected are reported. The fraction $\kappa(t)$ plays a crucial role for taking into account the presence of asymptomatic, or however undetected, infectious cases, which is one of the distinguishing features of COVID-19. The idea of including the fraction was also mentioned in the literature (e.g., Piazzola, Tamellini and Tempone (2021) and Ansumali et al. (2020)), but a constant fraction was often considered, which was criticized by Ansumali et al. (2020) as unrealistic. A dynamic fraction as in (3) varying with time is more realistic for the problem. It should be noted that in the model (3), it is intrinsically assumed that the rate of transmission by an infected-asymptomatic person is the same as the infected-symptomatic person. More sophisticated

models with an additional "asymptomatic-but-infected" compartment, which allow for different transmission rates in the model, may be considered, such as Robinson and Stilianakis (2013) and Ansumali et al. (2020). However, as pointed out by Ansumali et al. (2020), several recent studies show that there is no discernible difference between the two rates in the COVID-19 outbreak. See, for example, He et al. (2020); Li et al. (2020); Liu et al. (2020); Wölfel et al. (2020). As such, equal transmission rates are assumed in the proposed model, and the extension to the compartmental models with the asymptomatic-but-infected compartment, such as the SAIR model of Ansumali et al. (2020), is left for the future work. Such work could also allow for studying whether the effects of potential factors differ among the asymptomatic versus symptomatic persons, such as government interventions.

Notably, as pointed out by Ansumali et al. (2020), another popular epidemic model for the outbreak, SEIR model (Susceptible-Exposed-Infectious-Removed) (Kermack and McKendrick, 1927), is not as realistic, because the exposed group (E) of the SEIR model does not infect the susceptible group (S) as the E group does not carry a sufficient viral load to infect others through contact. This is unlike the asymptomatic-but-infected individuals in the COVID-19 outbreak, which *do* lead to the S group getting infected. As such, a model like the proposed model accounting for these asymptomatic individuals is more realistic.

The functional parameters in the SIR model are assumed to follow a joint Gaussian process (GP) prior:

$$\text{logit}\begin{pmatrix}\beta(\cdot)\\ \gamma(\cdot)\end{pmatrix} \sim \mathcal{GP}\left(\begin{bmatrix}\mu_1(\cdot)\\ \mu_2(\cdot)\end{bmatrix}, \tau\mathbf{A}\begin{bmatrix}K_{\boldsymbol{\phi}_1}(\cdot,\cdot) & 0\\ 0 & K_{\boldsymbol{\phi}_2}(\cdot,\cdot)\end{bmatrix}\mathbf{A}\right), \tag{4}$$

where $\text{logit}(x) = \log\frac{x}{1-x}$ and

$$\mathbf{A} = \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}^{1/2} = \frac{1}{\sqrt{2+2\sqrt{1-\rho^2}}}\begin{bmatrix}1+\sqrt{1-\rho^2} & \rho\\ \rho & 1+\sqrt{1-\rho^2}\end{bmatrix}. \tag{5}$$

The logit transformation is used here because both $\beta$ and $\gamma$ are rates which are bounded from zero to one, but the GP prior has positive measures over the negative reals. Other transformation, such as the probit function, $\Phi^{-1}(x)$, the cumulative log-log function, $\log(-\log(x))$, or the identity function, $x$, could be also used here. $\mu_j(\cdot)$ is the mean function, where we assume a constant mean, i.e., $\mu_j(\mathbf{x}) = \mu_j$. $\tau > 0$ is the process variance, and $K_{\boldsymbol{\phi}_j}$ is the correlation function, for which a Gaussian correlation function is commonly used in the form of $K_{\boldsymbol{\phi}_j}(\mathbf{x},\mathbf{x}') = \exp(-\|\boldsymbol{\phi}_j \odot (\mathbf{x}-\mathbf{x}')\|_2^2)$ for any $\mathbf{x},\mathbf{x}' \in \Omega \subseteq \mathbb{R}^d$, where $\boldsymbol{\phi}_j \in \mathbb{R}^d$ is the unknown lengthscale parameter and $\odot$ denotes the element-wise product of two vectors. Note that the correlation function is usually reparameterized as

$$K_{\boldsymbol{\phi}_j}(\mathbf{x},\mathbf{x}') = \prod_{l=1}^{d} \phi_{jl}^{4(x_l-x_l')^2} \quad \text{for any} \quad \mathbf{x},\mathbf{x}' \in \Omega, \tag{6}$$

where $\boldsymbol{\phi}_j = (\phi_{j1},\ldots,\phi_{jd}) \in (0,1)^d$, for the purpose of numerical stability, because the domain of $\phi_{jl} \in (0,1)$ is now bounded. See, for example, Brown and Atamturktur (2018) and Mak et al. (2018). As a result, the form of the kernel function (6) is used throughout this article.

In (4), we assume that $\text{logit}(\beta(\cdot))$ and $\text{logit}(\gamma(\cdot))$ are correlated with a positive *cross-correlation*, $0 < \rho \leq 1$, which implies that for any given $\mathbf{x} \in \Omega$, the correlation between $\text{logit}(\beta(\mathbf{x}))$ and $\text{logit}(\gamma(\mathbf{x}))$ is $\rho$. This can be verified by (5) and the fact that $K_{\boldsymbol{\phi}_1}(\mathbf{x},\mathbf{x}) = K_{\boldsymbol{\phi}_2}(\mathbf{x},\mathbf{x}) = 1$ for any $\mathbf{x} \in \Omega$. The dependence assumption of the two parameters, $\beta(\cdot)$ and $\gamma(\cdot)$, is crucial and appealing from an epidemiological perspective. For the compartmental models like SIR, it is well known that the parameters are strongly *coupled* in the modeling literature. See, for example, the joint posterior distribution in Roda et al. (2020) which

shows that the two parameters in an SIR model are highly positively correlated. Thus, the independent GP assumption as in Brown and Atamturktur (2018) and Plumlee, Joseph and Yang (2016) is not valid in this application. Note that unlike the covariance structures in Banerjee and Gelfand (2002); Qian, Wu and Wu (2008) where the parameters $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are assumed to be identical, the covariance structure (4) adopts the one in Fricker, Oakley and Urban (2013) and Svenson and Santner (2016), which is more flexible as the two lengthscale parameters are not necessarily identical.

Lastly, the fraction $\kappa(t)$ is assumed to have a Gaussian process prior:

$$\operatorname{logit}\kappa(\cdot) \sim \mathcal{GP}(\mu_3, \nu K_\varphi(\cdot,\cdot)) \tag{7}$$

with $\nu > 0$ and $\varphi \in (0,1)$, where $K_\varphi$ has the same form of (6).

Suppose that we observe the reported infectious cases in $n$ days, which are denoted by $\mathbf{y}_n = (y_1, \ldots, y_n)$. Denote $\kappa_t = \kappa(t), \beta_t = \beta(\mathbf{x}_t)$, $\gamma_t = \gamma(\mathbf{x}_t)$, $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n), \boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$. Furthermore, denote $\mathbf{1}_n = (1, \ldots, 1)^T \in \mathbb{R}^{n\times 1}, \mathbf{K}_{\boldsymbol{\phi}_j} = (K_{\boldsymbol{\phi}_j}(\mathbf{x}_i, \mathbf{x}_k))_{1\le i,k\le n} \in \mathbb{R}^{n\times n}$, $\mathbf{K}_\varphi = (K_\varphi(i,k))_{1\le i,k\le n} \in \mathbb{R}^{n\times n}$, and denote $a_{ij}$ as the element $\mathbf{A}_{ij}$ . Then, together with the model assumptions (3), (4), (6) and (7), we have the following hierarchical model,

$$\begin{aligned}
y_t|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa} &\overset{\text{indep.}}{\sim} \text{Poi}(\kappa_t f(t,\beta_t,\gamma_t)) \quad \text{for} \quad t = 1,\ldots,n,\\
\operatorname{logit}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} &\sim \mathcal{N}_{2n}\left(\begin{bmatrix}\mu_1\mathbf{1}_n\\ \mu_2\mathbf{1}_n\end{bmatrix}, \tau\boldsymbol{\Sigma}\right),\\
\operatorname{logit}(\boldsymbol{\kappa}) &\sim \mathcal{N}_n(\mu_3\mathbf{1}_n, \nu\mathbf{K}_\varphi),\\
\tau &\sim \text{InvGamma}(a_\tau, b_\tau), \qquad (8)\\
\rho &\sim \text{Beta}(1, b_\rho), \qquad (9)\\
\nu &\sim \text{InvGamma}(a_\nu, b_\nu), \qquad (10)\\
\varphi &\sim \text{Beta}(1, b_\varphi), \qquad (11)\\
\phi_{jl} &\overset{\text{indep.}}{\sim} \text{Beta}(1, b_\phi) \quad \text{for} \quad j = 1,2; l = 1,\ldots,d, \qquad (12)\\
\mu_j &\overset{\text{indep.}}{\sim} \mathcal{N}(\alpha_j, \sigma_j^2) \quad \text{for} \quad j = 1,2,3, \qquad (13)
\end{aligned}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} a_{11}^2\mathbf{K}_{\boldsymbol{\phi}_1} + a_{12}^2\mathbf{K}_{\boldsymbol{\phi}_2} & a_{11}a_{21}\mathbf{K}_{\boldsymbol{\phi}_1} + a_{12}a_{22}\mathbf{K}_{\boldsymbol{\phi}_2} \\ a_{11}a_{21}\mathbf{K}_{\boldsymbol{\phi}_1} + a_{12}a_{22}\mathbf{K}_{\boldsymbol{\phi}_2} & a_{21}^2\mathbf{K}_{\boldsymbol{\phi}_1} + a_{22}^2\mathbf{K}_{\boldsymbol{\phi}_2} \end{bmatrix},$$

and (8), (9), (10), (11), (12), (13) are the priors of the parameters, in which $\text{InvGamma}(a,b)$ is an inverse gamma distribution with shape parameter $a$ and rate parameter $b$, and $\text{Beta}(1,b)$ is a beta distribution with parameters 1 and $b$.

3.2. *Posterior Distributions.* The goal of this study is to infer the functional parameters $\beta(\mathbf{x})$ and $\gamma(\mathbf{x})$ and subsequently investigate whether the $d$-dimensional factor $\mathbf{x}$ plays a role in varying the basic reproduction number, which is denoted by $\mathcal{R}_0(\mathbf{x}) := \beta(\mathbf{x})/\gamma(\mathbf{x})$. In addition, predicting the number of future infections based on forecast weather and government interventions, say $\mathbf{x}_{n+1}$, is also of great interest. Therefore, the joint posterior distribution of $\beta(\mathbf{x}), \gamma(\mathbf{x})$, and the number of the future daily infected cases are developed as follows.

We first derive the joint posterior distribution of $\beta(\mathbf{x})$ and $\gamma(\mathbf{x})$. Denote the parameters $\boldsymbol{\psi} = (\tau, \rho, \nu, \varphi, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \mu_1, \mu_2)$ and data $= \{y_t, \mathbf{x}_t\}_{1\le t\le n}$. Then, the posterior distribution

given observations can be obtained by

$$\pi(\beta(\mathbf{x}),\gamma(\mathbf{x}),\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\text{data}) \propto \pi(\beta(\mathbf{x}),\gamma(\mathbf{x})|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi},\text{data})\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\text{data}),$$

where $\pi(x|y)$ denotes the posterior distribution of $x$ given $y$. Thus, the joint posterior distribution of $\beta(\mathbf{x})$ and $\gamma(\mathbf{x})$ can be approximated by Markov chain Monte Carlo (MCMC) by drawing the samples from $\pi(\beta(\mathbf{x}),\gamma(\mathbf{x})|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi},\text{data})$ and $\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\text{data})$ iteratively. The posterior $\pi(\beta(\mathbf{x}),\gamma(\mathbf{x})|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi},\text{data})$ can be drawn based on the property of conditional multivariate normal distributions, that is,

$$\text{logit}\begin{pmatrix}\beta(\mathbf{x})\\ \gamma(\mathbf{x})\end{pmatrix}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi},\text{data} \sim \mathcal{N}_2\left(u(\mathbf{x}),s(\mathbf{x})\right), \tag{14}$$

where

$$u(\mathbf{x}) = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} + \Sigma(\mathbf{x})\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\text{logit}(\boldsymbol{\beta}) - \mu_1\mathbf{1}_n\\ \text{logit}(\boldsymbol{\gamma}) - \mu_2\mathbf{1}_n\end{pmatrix}$$

and

$$s(\mathbf{x}) = \tau\left(\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix} - \Sigma(\mathbf{x})\boldsymbol{\Sigma}^{-1}\Sigma(\mathbf{x})^T\right)$$

with

$$\Sigma(\mathbf{x}) = \begin{bmatrix} a_{11}^2\mathbf{k}_{\boldsymbol{\phi}_1}(\mathbf{x}) + a_{12}^2\mathbf{k}_{\boldsymbol{\phi}_2}(\mathbf{x}) & a_{11}a_{21}\mathbf{k}_{\boldsymbol{\phi}_1}(\mathbf{x}) + a_{12}a_{22}\mathbf{k}_{\boldsymbol{\phi}_2}(\mathbf{x})\\ a_{11}a_{21}\mathbf{k}_{\boldsymbol{\phi}_1}(\mathbf{x}) + a_{12}a_{22}\mathbf{k}_{\boldsymbol{\phi}_2}(\mathbf{x}) & a_{21}^2\mathbf{k}_{\boldsymbol{\phi}_1}(\mathbf{x}) + a_{22}^2\mathbf{k}_{\boldsymbol{\phi}_2}(\mathbf{x})\end{bmatrix},$$

where $\mathbf{k}_{\boldsymbol{\phi}_j}(\mathbf{x}) = (K_{\boldsymbol{\phi}_j}(\mathbf{x},\mathbf{x}_1),\ldots,K_{\boldsymbol{\phi}_j}(\mathbf{x},\mathbf{x}_n))$. The MCMC samples of $\beta(\mathbf{x})$ and $\gamma(\mathbf{x})$ can then be obtained by sampling from the multivariate normal distribution of (14) and taking the inverse of the logit function. For the posterior $\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\text{data})$, we have

$$\begin{aligned}
\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\text{data}) &\propto \pi(\text{data}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi})\pi(\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{\kappa},\boldsymbol{\psi})\pi(\boldsymbol{\kappa}|\boldsymbol{\psi})\pi(\boldsymbol{\psi})\\
&\propto \exp\left\{-\sum_{t=1}^{n}\kappa_t f(t,\beta_t,\gamma_t)\right\} \times \prod_{t=1}^{n}\kappa_t^{y_t} f(t,\beta_t,\gamma_t)^{y_t}\\
&\times \exp\left\{-\frac{1}{\tau}\begin{pmatrix}\text{logit}(\boldsymbol{\beta}) - \mu_1\mathbf{1}_n\\ \text{logit}(\boldsymbol{\gamma}) - \mu_2\mathbf{1}_n\end{pmatrix}^T\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\text{logit}(\boldsymbol{\beta}) - \mu_1\mathbf{1}_n\\ \text{logit}(\boldsymbol{\gamma}) - \mu_2\mathbf{1}_n\end{pmatrix}\right\}\\
&\times \exp\left\{-\frac{1}{\nu}\left(\text{logit}(\boldsymbol{\kappa}) - \mu_3\mathbf{1}_n\right)^T\mathbf{K}_\varphi^{-1}\left(\text{logit}(\boldsymbol{\kappa}) - \mu_3\mathbf{1}_n\right)\right\}\\
&\times |\boldsymbol{\Sigma}|^{-1}|\mathbf{K}_\varphi|^{-1}\tau^{n+a_\tau-1}\exp\{-b_\tau\tau\}(1-\rho)^{b_\rho-1}\nu^{n/2+a_\nu-1}\exp\{-b_\nu\nu\}\\
&\times \varphi^{b_\varphi-1}\prod_{j=1}^{2}\prod_{l=1}^{d}(1-\phi_{jl})^{b_\phi-1}\exp\left\{-\frac{1}{2}\sum_{j=1}^{2}\frac{(\mu_j-\alpha_j)^2}{\sigma_j^2}\right\}.
\end{aligned} \tag{15}$$

The samples from this posterior distribution can be drawn by Gibbs sampling with Metropolis-Hastings algorithm, the details of which are given in Section 1 of the Supplementary Material (Sung, 2022).

Now we move to the posterior distribution of the number of future infections. Let $\mathbf{x}_{n+1}$ be the forecast weather and government interventions at time $n+1$, and denote $\beta_{n+1} =$

$\beta(\mathbf{x}_{n+1}), \gamma_{n+1} = \gamma(\mathbf{x}_{n+1}), \kappa_{n+1} = \kappa(n+1)$. Then the posterior distribution of the infected number, $y_{n+1}$, given the current observed data has

$$\begin{aligned}\pi(y_{n+1},&\beta_{n+1},\gamma_{n+1},\kappa_{n+1},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\mathbf{x}_{n+1},\text{data})\\ &\propto\pi(y_{n+1}|\beta_{n+1},\gamma_{n+1},\kappa_{n+1})\pi(\beta_{n+1},\gamma_{n+1}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi},\mathbf{x}_{n+1})\\ &\times\pi(\kappa_{n+1}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi})\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\text{data}).\end{aligned}$$

Similarly, this posterior distribution can be drawn by MCMC sampling, where the samples for $\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi}|\text{data})$ can be drawn as introduced before, and the samples from $\pi(\beta_{n+1},\gamma_{n+1}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi},\mathbf{x}_{n+1})$ can be similarly drawn from the multivariate normal distribution (14). The samples of $\kappa_{n+1}$ from $\pi(\kappa_{n+1}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi})$ can be drawn from the posterior of $\kappa(t)$,

$$\begin{aligned}\text{logit}\,(\kappa(t))|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\kappa},\boldsymbol{\psi},\sim\mathcal{N}(\mu_3+\mathbf{k}_\varphi(t)\mathbf{K}_\varphi^{-1}(\text{logit}\,(\boldsymbol{\kappa})-\mu_3\mathbf{1}_n),\\ \nu(1-\mathbf{k}_\varphi(t)\mathbf{K}_\varphi^{-1}\mathbf{k}_\varphi(t)^T)),\end{aligned} \tag{16}$$

and set $t = n+1$, where $\mathbf{k}_\varphi(t) = (K_\varphi(t,1),\ldots,K_\varphi(t,n))$. The distribution, $\pi(y_{n+1}|\beta_{n+1},\gamma_{n+1},\kappa_{n+1})$, follows a Poisson distribution with the mean $\kappa_{n+1}f(n+1,\beta_{n+1},\gamma_{n+1})$. Thus, the MCMC samples can be drawn iteratively from these posteriors.

**4. Simulation Study.** In this section, simulation studies are conducted to examine the performance of the proposed method. In the simulations, the hyperparameters in the priors (8), (9), (10), (11), (12), (13), are set as follow. Similar to Brown and Atamturktur (2018), the shape parameters $b_\rho, b_\phi$ and $b_\varphi$ are chosen to be 0.1, which place most probability mass near one to enforce the smoothness for the functional parameters; $a_\tau = a_\nu = 0.01$ and $b_\tau = b_\nu = 0.01$ are chosen so that the prior is centered at one with standard deviation $\sqrt{0.01/0.01^2} = 10$; for (13) we set $\alpha_j = 0$ and $\sigma_j^2 = 1$ for $j = 1,2,3$. For the MCMC sampling, 2,000 iterations are performed in a burn-in period, and after that additional 2,000 MCMC samples are drawn, which are thinned to reduce autocorrelation.

Suppose that the observation $y_t$ is simulated from a Poisson distribution with the mean function, $\kappa(t)f(t,\beta(x),\gamma(x))$, where $f(t,\beta(x),\gamma(x)) = (t/10+5\beta(x)+\gamma(x)(t/10)^2)^2$ and $x$ is one-dimensional factor in the space $[0,1]$. Let $\beta(x) = \sin(3x)\exp(-x)+0.2, \gamma(x) = \sin(3x)$, and $\kappa(t) = \exp(-t/50)$, which are demonstrated in the left and middle panels of Figure 1. It can be seen that the two curves, $\beta(x)$ and $\gamma(x)$, share some similarity overall, which suggests that the dependence assumption of these two functions is necessary. We generate $x_1,\ldots,x_{80}$ from a uniform distribution, and randomly generate $n = 80$ observations, $y_1,\ldots,y_{80}$. The right panel of Figure 1 shows the random samples as dots, where the solid line is the true mean function, $\kappa(t)f(t,\beta(x),\gamma(x))$. We use the first 60 samples, $y_1,\ldots,y_{60}$, as the training dataset and the other 20 samples as the test dataset.

Figure 2 shows the posterior draws of $\beta(x), \gamma(x)$ and $\kappa(t)$. It can be seen that the posterior means can recover the true functions very well. The predictions on the test dataset are presented in Figure 3, which shows that the posterior mean is reasonably close to the true function. These results demonstrate that the proposed method can perform well for the models with functional parameters in terms of estimation and prediction.

**5. Application to COVID-19.** In this section, we use the proposed model to analyze the COVID-19 virus spread among the eight largest metropolitan areas in the United States (US). In particular, we use the model to estimate the impacts of weather and government interventions (and their interactions) on virus transmissibility, to forecast daily infected cases based on these factors, and to estimate the fraction of cases reported.

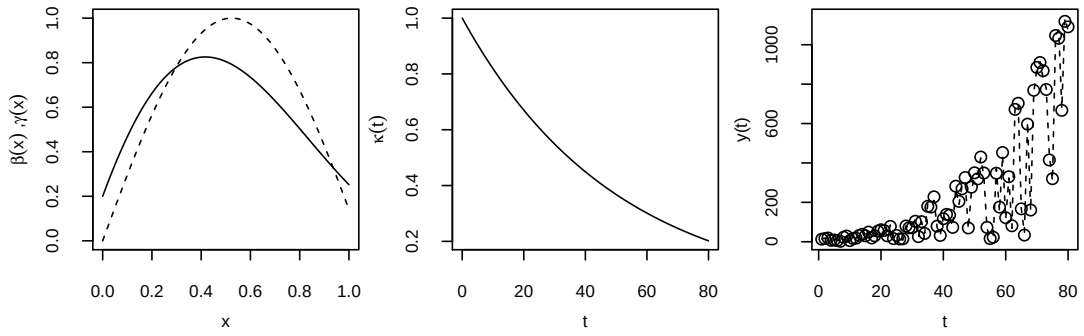


FIG 1. *Simulation setting. The left panel demonstrates $\beta(x)$ (solid line) and $\gamma(x)$ (dashed line), the middle panel demonstrates $\kappa(t)$, and the right panel demonstrates the true mean function, $\kappa(t)f(t,\beta(x),\gamma(x))$ (dashed line), and the simulated data (dots).*

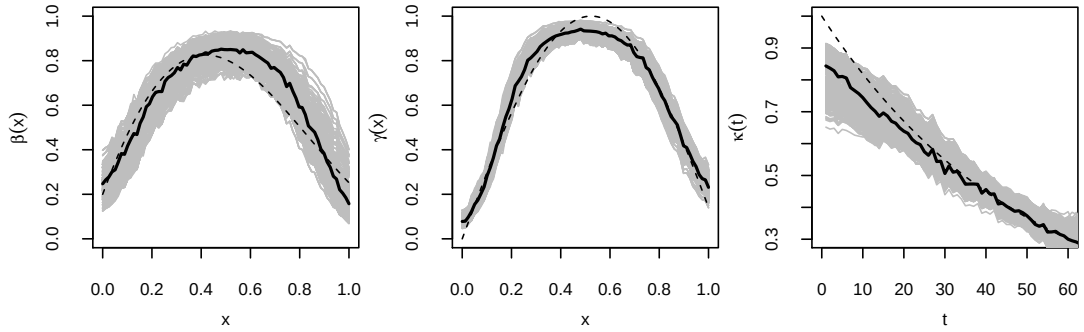


FIG 2. *Posteriors of $\beta(x)$ (left panel), $\gamma(x)$ (middle panel), and $\kappa(t)$ (right panel). The gray lines are the MCMC draws, the solid lines are the posterior mean, and the dashed lines are the true functions.*

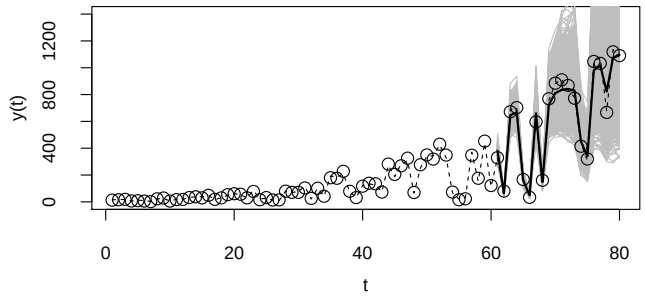


FIG 3. *Posteriors of test data $y_{61},\ldots,y_{80}$. The gray lines are the MCMC draws, the solid lines are the posterior mean, and the dashed lines are the true functions.*

The data source is briefly introduced here. The daily COVID-19 cases are obtained at the US county level from the data repository provided by New York Times (Almukhtar et al., 2020), starting from January 25, 2020 to November 25, 2020. The population sizes are obtained from the census bureau website, which also can be found in Yu (2020). The historical weather data and the weather forecast are collected from the Weather Underground (The Weather Company, 2020), which include the daily average temperature, humidity, wind

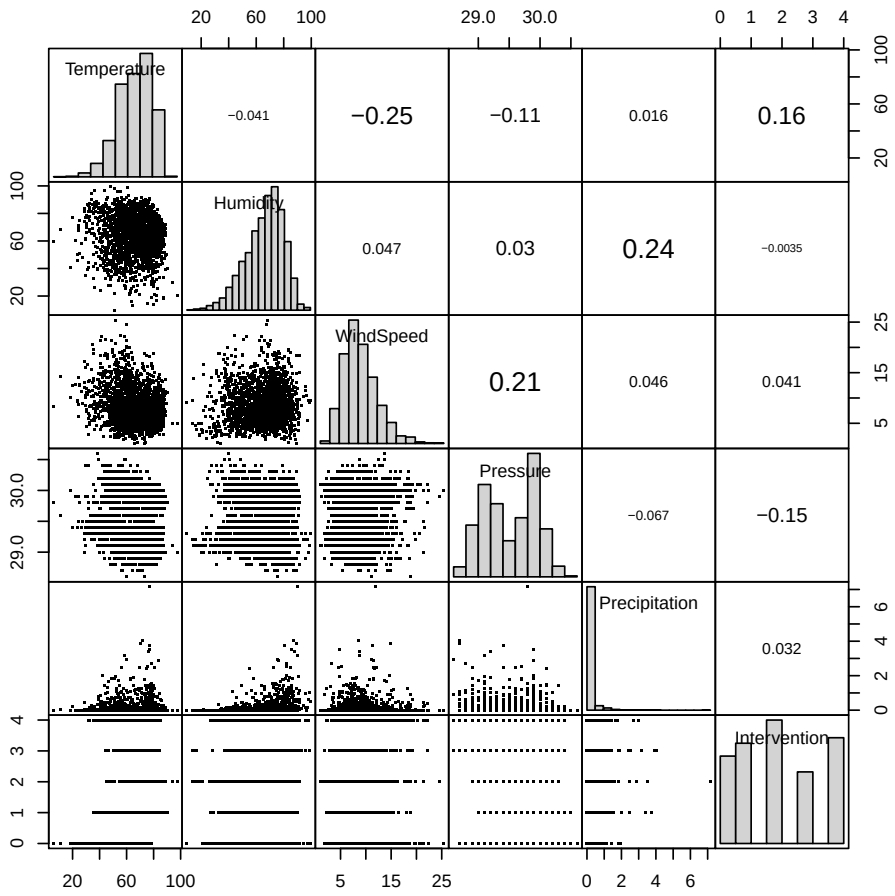


FIG 4. *Scatterplots of input factors, where the left diagonal panels are the scatterplots for every pair of factors, the right diagonal panels are the correlations, and the diagnoal panels are the histograms.*

speed, pressure, and precipitation. The information of government interventions is obtained from New York Times (Lee et al., 2020) and local media, where we categorize the interventions into five levels: (0) no intervention; (1) all businesses are open with mask required and some capacity limitations ; (2) all industries resume operations but some indoor services, such as bars and restaurants, remain closed; (3) Industries resume operations with severe restrictions and capacity limitations; (4) all non-essential businesses are closed. Scatterplots for every pair of factors are demonstrated in Figure 4, which appears to have no obvious relationship between any pair of the factors.

Now we are ready to apply the proposed model to the data, where the setting of the MCMC sampling is similar to the one in Section 4. Consider the confirmed cases before November 11 as the training data, and the cases from November 11 to 25 as the test data. Since the actual infectious period for COVID-19 is not available and it varies by individual and situation, as suggested by Centers for Disease Control and Prevention and Wilson (2020), we assume an infectious period of 11 days from the actual infection to the confirmation of the positive test result. In other words, we assume that the actual infection occurs 11 days prior to the confirmation date. The input factor is a 6-dimensional variable, i.e., $\mathbf{x} \in \mathbb{R}^6$, including 5 variables representing weather data and one variable representing government intervention levels. The MCMC samples of the basic reproduction number can be obtained from the MCMC samples

of $\beta(\mathbf{x})$ and $\gamma(\mathbf{x})$ by computing $\mathcal{R}_0(\mathbf{x}) = \beta(\mathbf{x})/\gamma(\mathbf{x})$. Since it's hard to visualize the function $\mathcal{R}_0(\mathbf{x})$ via a six-dimensional $\mathbf{x}$, similar to Welch et al. (1992), we use a functional ANOVA decomposition (Hoeffding, 1948; Sobol, 1993; Santner, Williams and Notz, 2018) for $\mathcal{R}_0(\mathbf{x})$ and plot its overall mean and main effects, which respectively measure the overall influence and the influence of a single factor on the basic reproduction number. That is, suppose that $\mathbf{x}$ follows a distribution $F(\mathbf{x})$ where $F(\mathbf{x}) = F_1(x_1) \times F_2(x_2) \times \cdots \times F_d(x_d)$, then the overall mean and the main effects of the function $\mathcal{R}_0(\mathbf{x})$ can be obtained by

$$m_0 := \int_{\Omega} \mathcal{R}_0(\mathbf{x}) \mathrm{d}F(\mathbf{x}) \quad \text{and} \quad m_j(x_j) = \int_{\Omega_{-j}} (\mathcal{R}_0(\mathbf{x}) - m_0) \mathrm{d}F_{-j}(\mathbf{x}_{-j}), \tag{17}$$

respectively, where $\int_{\Omega_{-j}} \cdots \mathrm{d}F_{-j}(\mathbf{x}_{-j})$ indicates integration over the variables not in $j$ and $F_{-j}(\mathbf{x}_{-j}) = \prod_{i \neq j}^{d} F_i(x_i)$. Since the MCMC samples of $\mathcal{R}_0(\mathbf{x})$ are available for any $\mathbf{x} \in \Omega$ and the integration in (17) can be approximated by the Monte-Carlo integration (Caflisch, 1998), the samples of the posterior distributions of $m_0$ and $m_j(x_j)$ can be naturally drawn via a Monte-Carlo sampling method. This is similar to Le Gratiet, Cannamela and Iooss (2014) for estimating the Sobol indices through a surrogate model that accounts for both the integration errors and the surrogate model uncertainty.

The boxplots of the overall means of $\mathcal{R}_0(\mathbf{x})$ are shown in Figure 5. It can be seen that among these eight cities, Chicago has the highest basic reproduction number than other cities, which implies that each existing infection in Chicago can cause more new infections than other cities, while the existing infection in Baltimore and Houston causes fewer new infections. Before illustrating the main effects, sensitivity analysis (Sobol, 1993) is adopted to determine which input factors are responsible for the most variation in the basic reproduction number. The result is shown in Figure 6. Although no unique factor can dominate the others for all of the cities in terms of sensitivity index, it appears that government interventions have made stronger impacts than weather factors on the virus spread in most of the cities, especially in Baltimore and San Francisco. On the other hand, some cities, such as Los Angeles, Saint Louis, and Atlanta, have shown evidence that temperature has played a crucial role in explaining the variation of the basic reproduction number.

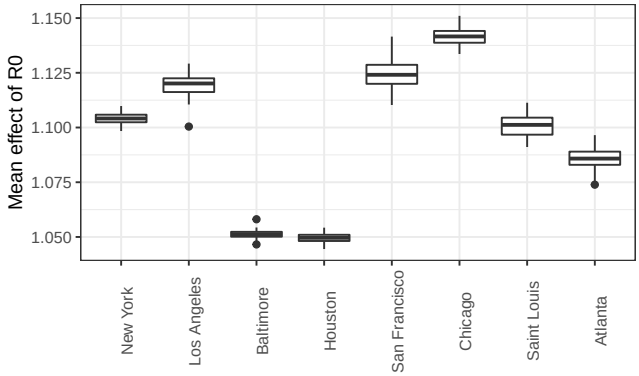


FIG 5. *Overall mean of basic reproduction number.*

The main effects of $\mathcal{R}_0(\mathbf{x})$ are demonstrated in Figure 7. As shown in the sensitivity analysis, the intervention and temperature factors both have larger variations in the main effects, ranging from -0.1 to 0.3, whereas the main effects of other weather factors mostly range from -0.1 to 0.1. Among these six factors, it shows that temperature and government intervention

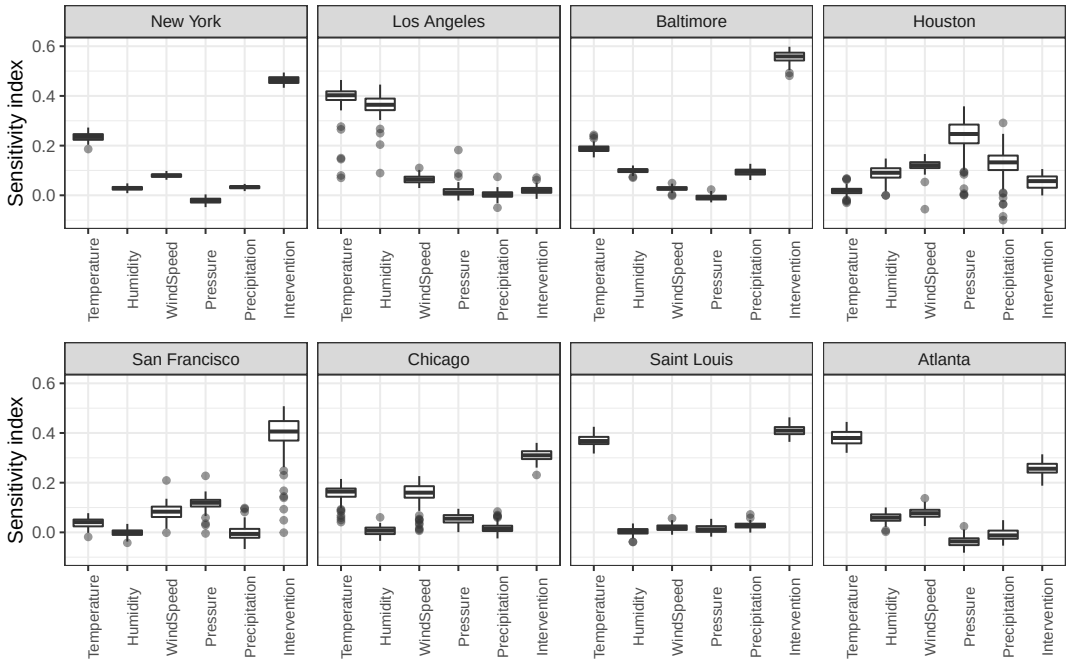


FIG 6. *Main effect indices of basic reproduction number.*

both have negative effects on the virus spread for all of the cities, while other factors have no common trend. In particular, it appears that a decrease of 10°F in temperature leads to an increase of roughly 0.06 (with a standard error (SE) of 0.0354) in the basic reproduction number. This result is quite promising in the sense that most of the existing methods cannot directly quantify the effect of temperature on the basic reproduction number. The intervention factor shows that the basic reproduction number can be effectively reduced if governments implement more restrictions to combat the COVID-19 outbreak, especially for New York and San Francisco, where a change from no intervention to the strictest restrictions can lead to a decrease in the basic reproduction number of approximately 0.42 (SE 0.03). This finding is consistent with the results in some recent work on the effect of government intervention for COVID-19, such as Flaxman et al. (2020); Haug et al. (2020); Haldar and Sethi (2020); Wang et al. (2020).

We further investigate the interaction effects of the basic reproduction number. Particularly, we focus on the interaction effects between the intervention factor, which is controllable by governments, and the five weather factors, which are uncontrollable. The sensitivity indices of these five interaction effects are first computed to compare their relative importance. For the sake of saving the space, only the interaction effects for New York are demonstrated here. The sensitivity indices and the interaction plot with the highest index are respectively shown in the left and right panels of Figure 8. It can be seen that the interaction effect between temperature and government interventions has the highest sensitivity index, and from the interaction plot of the two factors, it appears that when governments implement more restrictions, the effect of temperature on the virus spread tends to be milder. This result suggests that as the weather gets colder, policymakers may need to implement more restrictions to mitigate the pandemic.

We validate the proposed model by performing predictions on the test data from November 12 to 25. The prediction results of the eight cities are shown in Figure 9. The predictions are reasonably accurate over the 14-day period. Particularly, in the cities New York, Los Angeles,

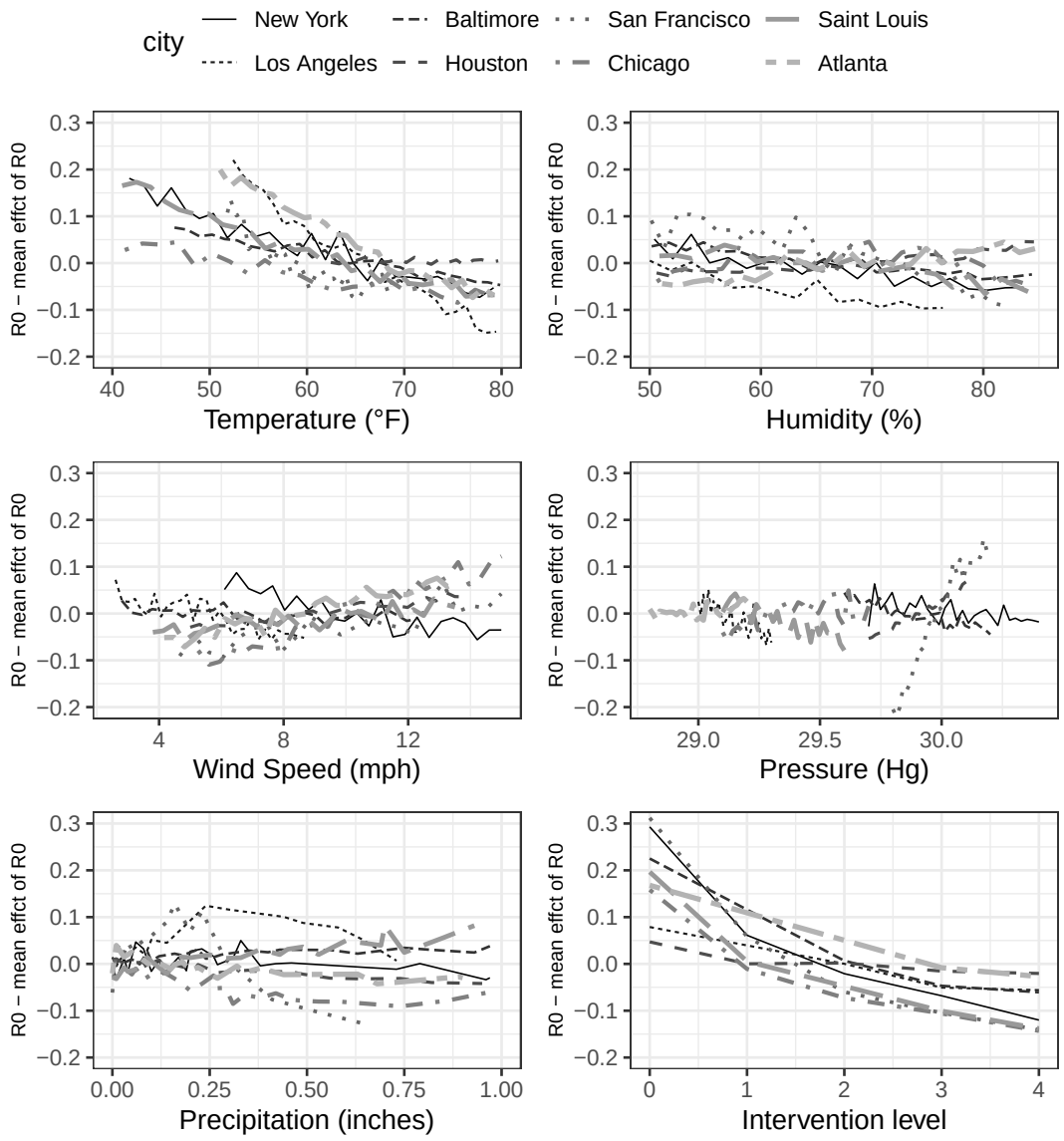


FIG 7. *Main effect of basic reproduction number, $m_j(x_j)$, which illustrates the influence of a single factor on the reproduction number.*

Baltimore, and San Francisco, the infected cases tend to increase over the 14-day period and our predictions successfully capture the trend. This shows strong empirical justification for our model specification.

Figure 10 presents the posteriors of the fraction of cases reported, $\kappa(t)$. The posteriors show that the actual infections are greatly undetected in most of the cities. This finding coincides with the recent results by U.S. Centers for Disease Control and Prevention (2021); Pei et al. (2021); Noh and Danuser (2021). In particular, San Francisco shows that the fraction less than 40% for the entire time of the ongoing pandemic, and New York shows low fractions during the peak of confirmed cases, which suggests that the number of actual infections during the peak is likely much higher (over 20,000 daily cases) than the confirmed daily cases (about 8,000). The estimation of the fractions provides crucial insights for pub-

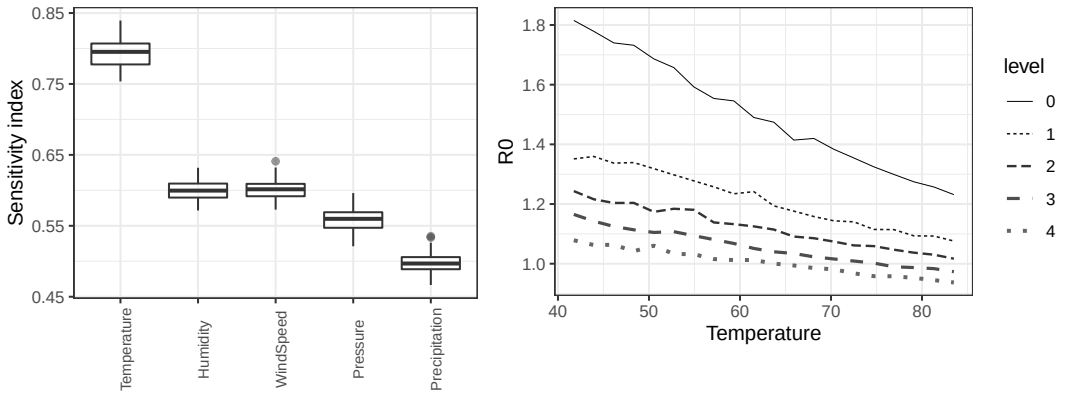


FIG 8. *Interaction effect indices (left) and interaction plot (right) between temperature and government intervention for New York.*

lic health, which determines the actual severity of COVID-19 and can be used to develop effective strategies against the outbreak.

To examine the robustness of the results, sensitivity analysis for the priors is conducted, which examines the impact of 30 different prior specifications on the resulting posteriors of $\mathcal{R}_0(\mathbf{x})$ for the test data. The percentage deviations in the average posterior estimates between the original model and the models with the 30 alternative prior specifications are computed, which are presented in Figure 11. It appears that the results are somewhat stable under different prior settings, with the average percentage deviation being about 3.6%. The posteriors based on the data of New York city are found to have higher shifts with some prior settings, indicating that it may require a bit more care in the future analysis.

**6. Concluding Remarks.** How the weather and government interventions affect the spread of a disease has been an important question but remains unclear in the literature. A new statistical model incorporated with the prominent SIR model is employed to study the impact on the COVID-19 virus transmissibility among eight US metropolitan areas. Gaussian process modeling and sensitivity analysis for the functional parameters enable to investigate the the main and interaction effects of the factors, which could lead to a new intervention strategy for policymakers. This study shows that, among six factors, temperature and government interventions have stronger impacts on the COVID-19 spread in most of the cities. The temperature has been found to have negative effects in all of the cities. Other weather factors, such wind speed and pressure, do not show common effects among the eight cities. New York city has shown a strong interaction effect between temperature and interventions, which suggests that more restrictions are necessary to mitigate the outbreak as the weather gets colder.

Although we found some potential associations between weather and virus transmissibility, it is worth emphasizing that these associations may not directly imply the causation of the virus transmissibility, meaning that there might be some lurking/causal variables which are correlated with these factors that make the associations appear stronger. For instance, as recent studies have shown (e.g., Wilson (2020); Soucy et al. (2020)), the individual mobility may have the direct impact on the COVID-19 spread, which could be strongly correlated with weather factors. Therefore, incorporating the information of individual mobility and estimating the causal effects of mobility and weather is worthwhile to investigate in the future work. In addition, it is conceivable to consider other potential factors for the virus transmission, such as private sector (i.e., non-governmental) interventions, travel restrictions, and public

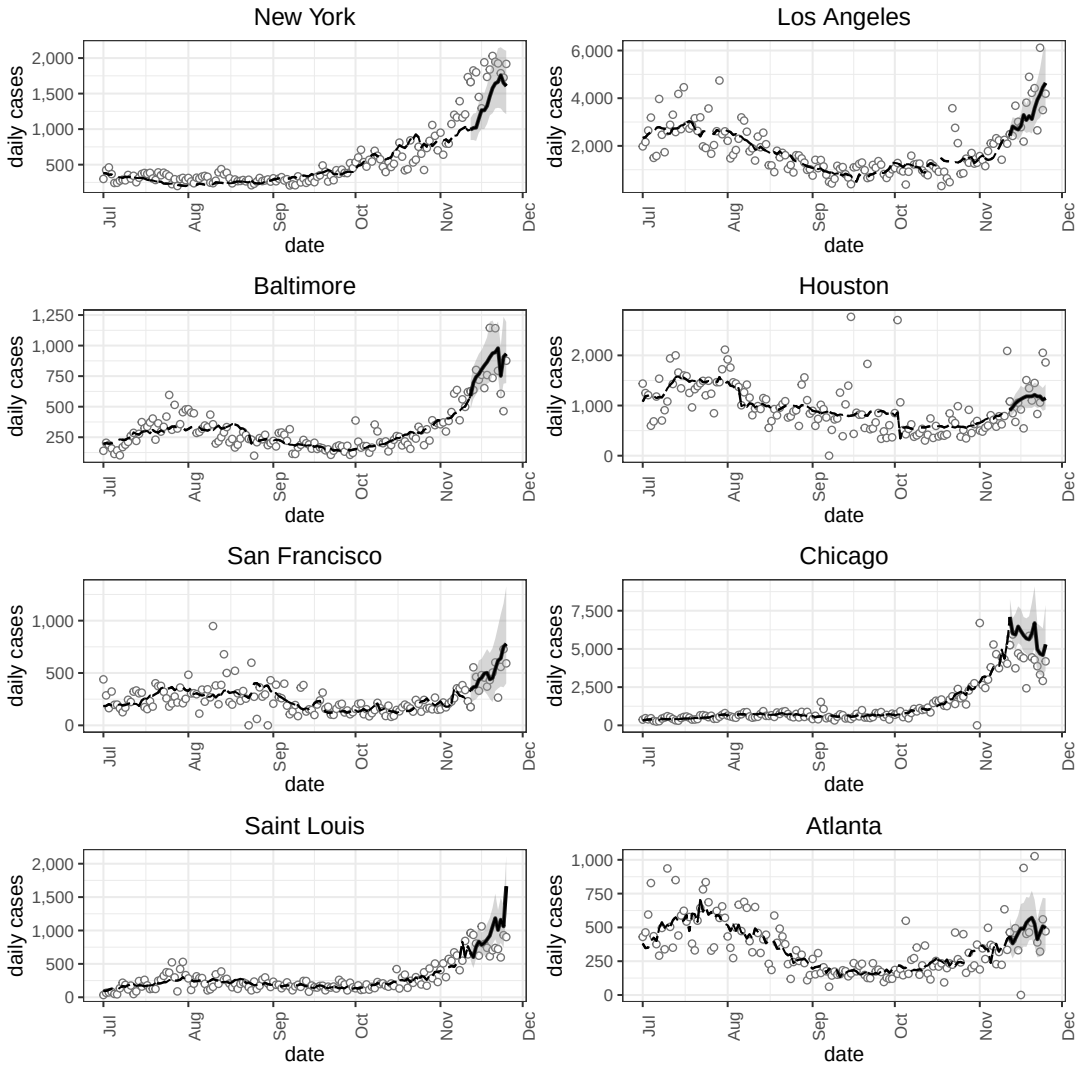


FIG 9. *Prediction performance of the proposed model. The dots are the confirmed numbers, the dashed lines are the fitted values from July 1 to November 11, the gray lines and the solid lines are the MCMC draws and posterior means for the test data from November 12 to 25.*

compliance with government recommendations such as vaccinations, quarantines and face coverings, if the data are available. However, including too many factors may lead to over-parameterization which in turn causes unstable results. A potential solution is to perform sensitivity analysis to screen out non-influential factors that have the least effect, and then remove them from the analysis. We explore these issues in future work.

**Acknowledgments.** The author gratefully acknowledges the conscientious efforts of the editor and two anonymous reviewers whose comments greatly strengthen this paper. The author is grateful to Dr. Ying Hung for the initial discussion which inspires the author to work on this problem.

**Funding.** The author was supported by NSF Grant DMS-2113407.

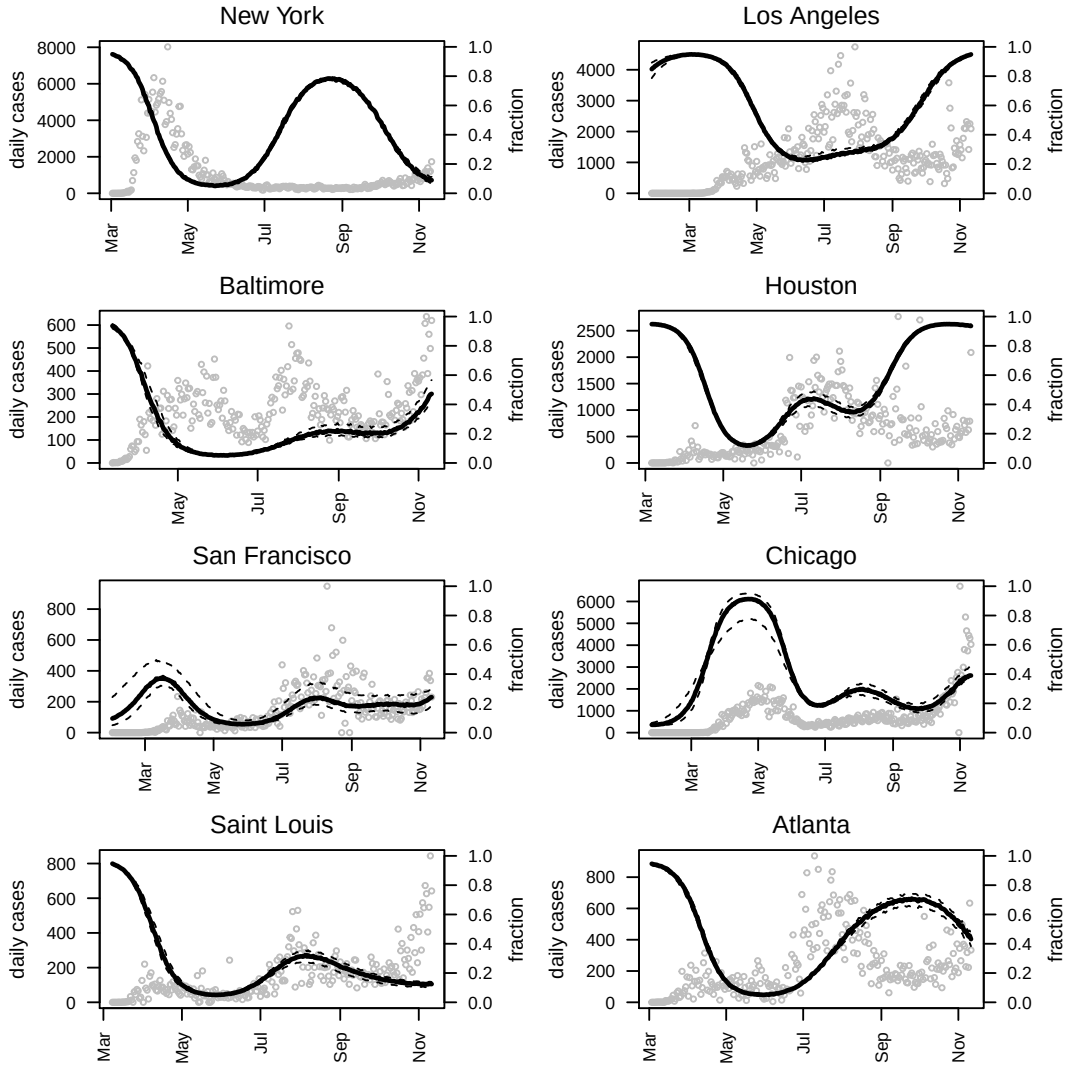


FIG 10. *The posteriors of the fraction of cases reported, $\kappa(t)$, where the solid and dashed lines are the median estimate and 95% confidence intervals, respectively, and the gray dots are the daily confirmed cases.*

## SUPPLEMENTARY MATERIAL

**Sampling details for the posterior distributions**

The details of sampling for the posterior distributions in Section 3.2 are described in this file.

**Data and R code**

A zip file containing the data and R code for reproducing the results in Sections 4 and 5.

## REFERENCES

ALMUKHTAR, S., AUFRICHTIG, A., BARNARD, A., BLOCH, M., CAI, W., CALDERONE, J., COLLINS, K., CONLEN, M., COOK, L., GIANORDOLI, G., HARMON, A., HARRIS, R., HASSAN, A., HUANG, J., ISSAWI, D., IVORY, D., LAI, K. K. R., LEMONIDES, A., MCCANN, A., OPPEL JR., R. A., PATEL, J. K., SEMPLE, K., SHAVER, J. W., SINGHVI, A., SMART, C., SMITH, M., SUN, A., WATKINS, D., WILLIAMS, T., WU, J. and YOURISH, K. (2020). Coronavirus in the US: Latest map and case count. *The New York Times*.

AMBROSIO, B. and AZIZ-ALAOUI, M. (2020). On a coupled time-dependent SIR models fitting with New York and New-Jersey states COVID-19 data. *Biology (Basel)* **9** 135.

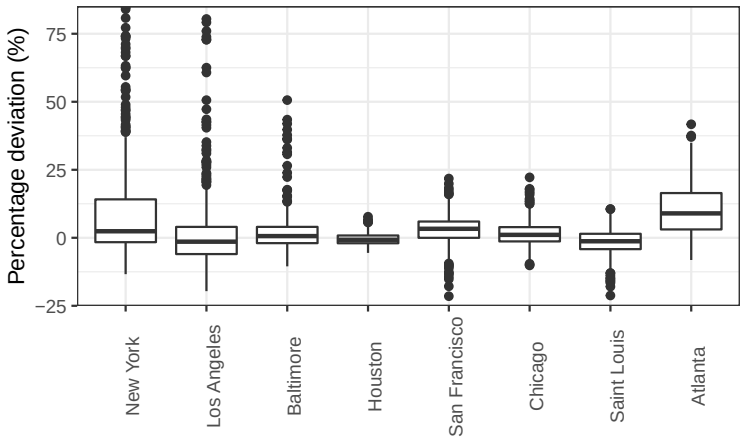


FIG 11. *The sensitivity analysis for the priors across 30 different prior specifications, which shows the percentage deviations between the original posteriors and the posteriors with the alternative priors.*

ANSUMALI, S., KAUSHAL, S., KUMAR, A., PRAKASH, M. K. and VIDYASAGAR, M. (2020). Modelling a pandemic with asymptomatic patients, impact of lockdown and herd immunity, with applications to SARS-CoV-2. *Annual Reviews in Control* **50** 432–447.

BANERJEE, S. and GELFAND, A. E. (2002). Prediction, interpolation and regression for spatially misaligned data. *Sankhyā: The Indian Journal of Statistics, Series A* **64** 227–245.

BROWN, D. A. and ATAMTURKTUR, S. (2018). Nonparametric functional calibration of computer models. *Statistica Sinica* **28** 721–742.

CAFLISCH, R. E. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* **7** 1–49.

CARSON, R. T., CARSON, S. L., DYE, T. K., MAYFIELD, S. L., MOYER, D. C. and CHU, A. Y. (2020). COVID-19's US Temperature Response Profile. *medRxiv*.

CAUCHEMEZ, S., NOUVELLET, P., CORI, A., JOMBART, T., GARSKE, T., CLAPHAM, H., MOORE, S., MILLS, H. L., SALJE, H., COLLINS, C., RODRIQUEZ-BARRAQUER, I., RILEY, S., TRUELOVE, S., ALGARNI, H., ALHAKEEM, R., ALHARBI, K., TURKISTANI, A., AGUAS, R. J., CUMMINGS, D. A., VAN KERKHOVE, M. D., DONNELLY, C. A., LESSLER, J., FRASER, C., AL-BARRAK, A. and FERGUSON, N. M. (2016). Unraveling the drivers of MERS-CoV transmission. *Proceedings of the National Academy of Sciences* **113** 9081–9086.

CHEN, Y.-C., LU, P.-E., CHANG, C.-S. and LIU, T.-H. (2020). A Time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE Transactions on Network Science and Engineering*.

THE WEATHER COMPANY (2020). Weather Underground. https://www.wunderground.com/.

COOPER, I., MONDAL, A. and ANTONOPOULOS, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals* **139** 110057.

COWLING, B. J., HO, L. M. and LEUNG, G. M. (2008). Effectiveness of control measures during the SARS epidemic in Beijing: a comparison of the Rt curve and the epidemic curve. *Epidemiology and Infection* **136** 562–566.

COWLING, B. J., ALI, S. T., NG, T. W., TSANG, T. K., LI, J. C., FONG, M. W., LIAO, Q., KWAN, M. Y., LEE, S. L., CHIU, S. S., PROF WU, J. T., WU, P. and LEUNG, G. M. (2020). Impact assessment of non-pharmaceutical interventions against COVID-19 and influenza in Hong Kong: an observational study. *Lancet Public Health* **5** e279–e288.

D'ARIENZO, M. and CONIGLIO, A. (2020). Assessment of the SARS-CoV-2 basic reproduction number, R0, based on the early phase of COVID-19 outbreak in Italy. *Biosafety and Health*.

DIEKMANN, O., HEESTERBEEK, H. and BRITTON, T. (2012). *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press.

DIETZ, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research* **2** 23–41.

FLAXMAN, S., MISHRA, S., GANDY, A., UNWIN, H. J. T., MELLAN, T. A., COUPLAND, H., WHITTAKER, C., ZHU, H., BERAH, T., EATON, J. W., MONOD, M., IMPERIAL COLLEGE COVID-19 RESPONSE TEAM, GHANI, A. C., DONNELLY, C. A., RILEY, S., VOLLMER, M. A. C., FERGUSON, N. M., OKELL, L. C. and BHATT, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584** 257–261.

U. S. CENTERS FOR DISEASE CONTROL AND PREVENTION (2021). Estimated disease burden of COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/.

FRICKER, T. E., OAKLEY, J. E. and URBAN, N. M. (2013). Multivariate Gaussian process emulators with nonseparable covariance structures. *Technometrics* **55** 47–56.

GUPTA, A., PRADHAN, B. and MAULUD, K. N. A. (2020). Estimating the impact of daily weather on the temporal pattern of COVID-19 outbreak in India. *Earth Systems and Environment* **4** 523–534.

HALDAR, A. and SETHI, N. (2020). The effect of country-level factors and government intervention on the incidence of COVID-19. *Asian Economics Letters* **1** 17804.

HAUG, N., GEYRHOFER, L., LONDEI, A., DERVIC, E., DESVARS-LARRIVE, A., LORETO, V., PINIOR, B., THURNER, S. and KLIMEK, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour* 1–10.

HE, X., LAU, E. H., WU, P., DENG, X., WANG, J., HAO, X., LAU, Y. C., WONG, J. Y., GUAN, Y., TAN, X., MO, X., CHEN, Y., LIAO, B., CHEN, W., HU, F., ZHANG, Q., ZHONG, M., WU, Y., ZHAO, L., ZHANG, F., COWLING, B. J., LI, F. and LEUNG, G. M. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26** 672–675.

HEESTERBEEK, H., ANDERSON, R. M., ANDREASEN, V., BANSAL, S., DE ANGELIS, D., DYE, C., EAMES, K. T., EDMUNDS, W. J., FROST, S. D., FUNK, S., HOLLINGSWORTH, T. D., HOUSE11, T., ISHAM, V., KLEPAC, P., LESSLER, J., LLOYD-SMITH, J. O., METCALF, C. J. E., MOLLISON, D., PELLIS, L., PULLIAM, J. R. C., ROBERTS, M. G., VIBOUD, C. and ISAAC NEWTON INSTITUTE IDD COLLABORATION (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347** aaa4339.

HOEFFDING, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *Annals of Mathematical Statistics* **19** 308-334.

HONG, H. G. and LI, Y. (2020). Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic. *PloS one* **15** e0236464.

JAMIL, T., ALAM, I., GOJOBORI, T. and DUARTE, C. M. (2020). No evidence for temperature-dependence of the COVID-19 epidemic. *Frontiers in Public Health* **8** 436.

KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B* **63** 425–464.

KERMACK, W. O. and MCKENDRICK, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **115** 700–721.

LE GRATIET, L., CANNAMELA, C. and IOOSS, B. (2014). A Bayesian approach for global sensitivity analysis of (multifidelity) computer codes. *SIAM/ASA Journal on Uncertainty Quantification* **2** 336–363.

LEE, J. C., MERVOSH, S., AVILA, Y., HARVEY, B., MATTHEWS, A. L., GAMIO, L., POPOVICH, N., MCCANN, A. and KEEFE, J. (2020). See coronavirus restrictions and mask mandates for all 50 states. *The New York Times*.

LI, R., PEI, S., CHEN, B., SONG, Y., ZHANG, T., YANG, W. and SHAMAN, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368** 489–493.

LIU, Y., YAN, L.-M., WAN, L., XIANG, T.-X., LE, A., LIU, J.-M., PEIRIS, M., POON, L. L. and ZHANG, W. (2020). Viral dynamics in mild and severe cases of COVID-19. *The Lancet Infectious Diseases* **20** 656–657.

MAK, S., SUNG, C.-L., WANG, X., YEH, S.-T., CHANG, Y.-H., JOSEPH, V. R., YANG, V. and WU, C. F. J. (2018). An efficient surrogate model for emulation and physics extraction of large eddy simulations. *Journal of the American Statistical Association* **113** 1443–1456.

NOH, J. and DANUSER, G. (2021). Estimation of the fraction of COVID-19 infected people in US states and countries worldwide. *PLOS ONE* **16** e0246772.

OSTHUS, D., HICKMANN, K. S., CARAGEA, P. C., HIGDON, D. and DEL VALLE, S. Y. (2017). Forecasting seasonal influenza with a state-space SIR model. *The Annals of Applied Statistics* **11** 202.

PEI, S., YAMANA, T. K., KANDULA, S., GALANTI, M. and SHAMAN, J. (2021). Burden and characteristics of COVID-19 in the United States during 2020. *Nature* to appear.

PIAZZOLA, C., TAMELLINI, L. and TEMPONE, R. (2021). A note on tools for prediction under uncertainty and identifiability of SIR-like dynamical systems for epidemiology. *Mathematical Biosciences* **332** 108514.

PLUMLEE, M., JOSEPH, V. R. and YANG, H. (2016). Calibrating functional parameters in the ion channel models of cardiac cells. *Journal of the American Statistical Association* **111** 500–509.

QIAN, P. Z. G., WU, H. and WU, C. F. J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **50** 383–396.

ROBINSON, M. and STILIANAKIS, N. I. (2013). A model for the emergence of drug resistance in the presence of asymptomatic infections. *Mathematical Biosciences* **243** 163–177.

RODA, W. C., VARUGHESE, M. B., HAN, D. and LI, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling* **5** 271-281.

SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2018). *The Design and Analysis of Computer Experiments*, Second ed. Springer New York.

SOBOL, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments* **1** 407–414.

SOUCY, J.-P. R., STURROCK, S. L., BERRY, I., DANEMAN, N., MACFADDEN, D. R. and BROWN, K. A. (2020). Estimating the effect of physical distancing on the COVID-19 pandemic using an urban mobility index. *medRxiv*.

SUNG, C. L. (2022). Supplement to "Estimating functional parameters for understanding the impact of weather and government iinterventions on COVID-19 outbreak". DOI:10.1214/[provided by typesetter].

SUNG, C.-L. and HUNG, Y. (2020). Efficient calibration for imperfect epidemic models with applications to the analysis of COVID-19. *arXiv preprint arXiv:2009.12523*.

SVENSON, J. and SANTNER, T. (2016). Multiobjective optimization of expensive-to-evaluate deterministic computer simulator models. *Computational Statistics & Data Analysis* **94** 250–264.

R CORE TEAM (2018). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

TUO, R. and WU, C. F. J. (2015). Efficient calibration for imperfect computer models. *The Annals of Statistics* **43** 2331–2352.

WANG, G., ZHANG, Y., ZHAO, J., ZHANG, J. and JIANG, F. (2020). Mitigate the effects of home confinement on children during the COVID-19 outbreak. *The Lancet* **395** 945–947.

WELCH, W. J., BUCK, R. J., SACKS, J., WYNN, H. P., MITCHELL, T. J. and MORRIS, M. D. (1992). Screening, predicting, and computer experiments. *Technometrics* **34** 15–25.

WILLIAMS, C. K. and RASMUSSEN, C. E. (2006). *Gaussian Processes for Machine Learning*. MIT press Cambridge, MA.

WILSON, D. J. (2020). Weather, social distancing, and the spread of COVID-19. *medRxiv*.

WÖLFEL, R., CORMAN, V. M., GUGGEMOS, W., SEILMAIER, M., ZANGE, S., MÜLLER, M. A., NIEMEYER, D., JONES, T. C., VOLLMAR, P., ROTHE, C., HOELSCHER, M., BLEICKER, T., BRÜNINK, S., SCHNEIDER, J., EHMANN, R., ZWIRGLMAIER, K., DROSTEN, C. and WENDTNER, C. (2020). Virological assessment of hospitalized patients with COVID-2019. *Nature* **581** 465–469.

XU, R., RAHMANDAD, H., GUPTA, M., DIGENNARO, C., GHAFFARZADEGAN, N., AMINI, H. and JALALI, M. S. (2020). The Modest Impact of Weather and Air Pollution on COVID-19 Transmission. *medRxiv*.

YU, X. (2020). Impact of mitigating interventions and temperature on the instantaneous reproduction number in the COVID-19 epidemic among 30 US metropolitan areas. *medRxiv*.

ZHANG, J., LITVINOVA, M., WANG, W., WANG, Y., DENG, X., CHEN, X., LI, M., ZHENG, W., YI, L., CHEN, X. et al. (2020). Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases* **20** 793–802.

ZHAO, S., LIN, Q., RAN, J., MUSA, S. S., YANG, G., WANG, W., LOU, Y., GAO, D., YANG, L., HE, D. and WANG, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases* **92** 214–217.