Statistical Inference Using GLEaM Model with Spatial Heterogeneity and Correlation between Regions

Yixuan Tan[1], Yuan Zhang[2], Xiuyuan Cheng[1], and Xiao-Hua Zhou[3,4,5]

[1]Department of Mathematics, Duke University
[2]School of Mathematical Sciences, Peking University
[3]Center for Statistical Sciences, Peking University
[4]Beijing International Center for Mathematical Research, Peking University
[5]Department of Biostatistics, School of Public Health, Peking University

Abstract

A better understanding of the various patterns in the coronavirus disease 2019 (COVID-19) spread in different parts of the world is crucial to its prevention and control. Motivated by the celebrated GLEaM model (Balcan et al., 2010[1]), this paper proposes a stochastic dynamic model to depict the evolution of COVID-19. The model allows spatial and temporal heterogeneity of transmission parameters and involves transportation between regions. Based on the proposed model, this paper also designs a two-step procedure for parameter inference, which utilizes the correlation between regions through a prior distribution that imposes graph Laplacian regularization on transmission parameters. Experiments on simulated data and real-world data in China and Europe indicate that the proposed model achieves higher accuracy in predicting the newly confirmed cases than baseline models.

1 Introduction

The outbreak of coronavirus disease 2019 (COVID-19) has impacted all aspects of the world significantly for a long. As of 26 Oct 2021, there have been over 243 million confirmed cases of COVID-19, including over 4 million deaths ([2]). Therefore, it is essential to study the spread of COVID-19 for better prediction and prevention of the disease. This paper first proposes a stochastic dynamical model that describes the spread of COVID-19 in multiple regions and then develops an algorithm to estimate the transmission parameters and their posterior distributions.

The model proposed in this paper is inspired by The Global Epidemic and Mobility (GLEaM) model proposed in [1]. GLEaM is a stochastic dynamic model to depict the spread of epidemics, integrating multiple data layers that couple metapopulations at the global level. The model involves 3362 subpopulations in 220 countries obtained from Voronoi tessellation, centered around major airports. These subpopulations are connected by a multi-layer mobility network composed of processes from short-range communication in geographically close subpopulations to international flights. In each subpopulation, the transmission of epidemics is modeled by a Susceptible-Exposed-Infected-Removed (SEIR) compartmental model that integrates short-range communication. After specifying the compartmental model and initial conditions, GLEaM can be run repeatedly to build a statistical ensemble for each point in the space of parameters, which can be further utilized for Monte Carlo analysis of epidemic parameters.

In the vast majority of GLEaM's applications ([3, 4, 5, 6, 7, 8, 9, 10]), the parameters are estimated based on [11]. [11] performed the maximal likelihood analysis of the reproduction number R_0 in the seed region Mexico. For each value of R_0, the method generated the distribution of arrival time of the influenza A(H1N1) in 12 countries produced by 2×10^3 GLEaM simulations. Then, the optimal R_0 was chosen by maximizing the likelihood function of arrival time. [11] and subsequent works following its settings ([3, 4, 5, 6, 7, 8, 9, 10]) assumed that the epidemic was seeded from one region and the transmission parameters or the other parameters (like the introduction date and location) were estimated through the maximal likelihood analysis of arrival time or other events. In particular, the method in [11] was adopted in [5] to estimate the posterior distribution of R_0 of COVID-19, which was assumed to be uniform for all subpopulations at all times. However, this setting is not suitable for the current scenario of the COVID-19 pandemic, since COVID-19 has lasted for a long time, and the community transmission has been widespread in most countries in the world [12]. To model the spread of COVID-19, both spatial and temporal heterogeneity of the transmission parameters are needed instead of directly modeling R_0 as a periodic function of time as in [11], since the social behaviors, containment measures, medical conditions, and other elements that might have effects on the spread of COVID-19 may vary among different countries and over time.

Recently, [13] improved the model described above by estimating the initially infected individuals in each subpopulation through microblogging data from Twitter and also estimating the reproduction number R_0 for USA, Italy, and Spain separately. Therefore, the method used in [13] involved spatial heterogeneity. However, in this study, international travel was not considered, and GLEaM was applied to each of the aforementioned countries (as an isolated system) independently. The transmission rates in all subpopulations of these countries were again presumed to be homogeneous. Furthermore, [13] also assumed that the initially infected individuals for each subpopulation were proportional to the total number of Twitter users in that subpopulation. Thus it still assumed that the severity of the pandemic at the initial outbreak of COVID-19 was uniform over the country, which is not the case for COVID-19.

In addition to the issues mentioned above, other potential concerns exist when using GLEaM to model the spread of COVID-19. As mentioned in the last section of [14], GLEaM can be used to simulate the spread of the epidemic under normal conditions since it uses the "steady-state" mobility data around the world. However, since the outbreak of COVID-19, the social order has been disrupted, and travel has been restricted in most countries. Thus GLEaM might not work well with its multi-layer mobility networks. Furthermore, the estimation of parameters using GLEaM is based on a large number of simulations to explore the space of parameters, which usually takes much computational time ([1]). In addition, although the social behavior, medical conditions, and other factors that affect the spread of COVID-19 may vary among different regions, these factors for regions that are geographically close or have similarities in other aspects still bear some resemblance. Hence, the transmission rates for COVID-19 should not only own heterogeneity but also correlate with each other. However, neither of the features is reflected in GLEaM or most of its applications.

As the consequences of the possible constraints of GLEaM described above, most of the papers using GLEaM to model the epidemics focus on estimating the transmission parameter in the "seed region" and mainly use the epidemic data at the beginning of the outbreak. However, since factors like government policies and social behavior result in spatial and temporal heterogeneity of the long-lasting spread of COVID-19 around the world, GLEaM and inference methods based on a large number of simulations using GLEaM may not be directly applied to estimate the transmission parameters of COVID-19 in different regions and periods.

In this paper, we propose a model inspired by the GLEaM framework that further enables both spatial and temporal heterogeneity of parameters, making it possible to model the ongoing spread of COVID-19 instead of only the initial outbreak as considered in most applications of GLEaM. Motivated by GLEaM,

we include inter-district mobility in the model when such mobility data are available. For the inference of the parameters, we introduce an optimization algorithm that utilizes the correlation between districts. Furthermore, the posterior distribution of parameters is estimated by Markov chain Monte Carlo (MCMC) sampling starting from an optimal choice of the parameters obtained from the optimization process. This takes a much shorter time than the methods based on GLEaM simulations.

The main contributions of our paper are:

- We propose a stochastic model to describe the epidemic's lasting spread, allowing spatial and temporal heterogeneity of transmission parameters and transportation between districts.
- Based on the model, we also design an algorithm that first makes inference for the parameters through a two-step procedure combining the information of correlation between districts, which is equivalent to imposing graph Laplacian regularization on the transmission parameters. Then, the algorithm estimates the posterior distribution efficiently by MCMC sampling with the estimated parameters as the initial points.
- We compare the performance of the proposed model with the baseline models on both simulated data and real-world data.

  - For the simulated data, the results show that combining heterogeneity and transportation into the model helps improve the performance of trajectory prediction and parameter estimation. Moreover, our inference algorithm that integrates the correlation of districts leads to further improvement in predicting the future trajectories.

  - For the real-world data in China and Europe, the proposed model still outperforms the baselines in trajectory prediction.

Data sets used in this paper are publicly available. Our work focuses on proposing a new method for parameter inference and, consequently, answering questions regarding the long-lasting spread of COVID-19 in multiple regions.

A strength of the proposed model resides in introducing spatial and temporal heterogeneity of transmission parameters. We remark that there exist recent works which also involved different levels of heterogeneity in their models in various ways. [15, 16] utilized randomness in reproduction number to reflect heterogeneity of the population, using plate model with Bayesian method and heterogeneous well-mixed theory [17] with age-of-infection method [18], respectively. In addition, [19, 20, 21, 22, 23] adapted SEIR / Susceptible-Exposed-Infected (SEI) / Susceptible-Infected (SI) compartmental models similar to this paper. Among these works, [19, 20, 21] considered heterogeneity in the aspects of age groups, social links, and vaccination status separately. [22, 23], including spatial heterogeneity in one county and among states respectively, bore more similarity with this paper since they also allowed transmission parameters to be spatially heterogeneous and involved transportation between different regions. However, [22] only considered intracounty data, and the transportation was used to compute the effective size of compartments and did not affect the dynamic model. Furthermore, the transmission rates in [22] were determined by an SDE whose parameters were to be fitted. Therefore, [22] focused on a different scope from this paper. The settings of compartments in [23] are more realistic than the one considered in this paper by considering reporting rates. Nevertheless, compared with the model and inference algorithm described in Section 2, transmission rates in [23] were kept constant over time and were estimated using independent non-informative prior. Both [22] and [23] used Ensemble Kalman Filter, which samples particles in the state space according to the prior distribution and obtains the posterior distribution in the process of moving particles at each time step. This may be computationally less efficient than directly applying

MCMC according to the posterior distribution with the initial point maximizing the likelihood function, as implemented in this paper.

The rest of the paper is structured as below. In Section 2, we introduce the stochastic dynamic model and corresponding inference algorithm. In Section 3, we compare the performance of trajectory prediction and parameter estimation of models with or without mobility, heterogeneity, or using correlation information in the inference part for the simulated data. Section 4 describes the real-world data used in this paper, presents the results and findings of applying our model to the COVID-19 data in China and Europe. We discuss our limitations and possible extensions in Section 5.

2 Method

2.1 Model Description

Motivated by the stochastic dynamic process constructed in [24], we propose a stochastic model that considers mobility.

Suppose there are n regions (region $1, ..., n$), and we consider the evolution of the epidemic within a period of T days (day $1, ..., T$). For each region, we consider the following epidemiology compartments:

- $S_k(t)$: Susceptible.
- $E_k(t)$: Exposed and infectious.
- $H_k(t)$: Hospitalized.
- $R_k(t)$: Recovered or dead.

The subscript k of the states denotes that they belong to the k-th region, and dependence on the continuous time t is addressed through expressing the states as functions of $t \in [0, T]$.

We also denote $N_k(t) = S_k(t) + E_k(t) + H_k(t) + R_k(t)$ as the total population in the k-th region at time t. $N_k(t)$ is allowed to be time-varying since the inter-region mobility is taken into consideration, especially for the days before the implementation of travel restrictions. In this report, we assume that $N = \sum_{k=1}^{n} N_k(t)$ keeps constant over time. We also denote $(N_a)_k(t) = S_k(t) + E_k(t) + R_k(t)$ as the total population that are permitted to move in the k-th region, excluding the hospitalized ones.

Furthermore, denote $W_{i,j}^t$ as the traveling volume from region i to j on the t-th day ($t = 1, \ldots, T$). Then the traveling matrix W_t on the t-th day can be written as

$$W_t = \begin{pmatrix} w_{1,1}^t & w_{1,2}^t & \cdots & w_{1,n}^t \\ w_{2,1}^t & w_{2,2}^t & \cdots & w_{2,n}^t \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1}^t & w_{n,2}^t & \cdots & w_{n,n}^t \end{pmatrix}. \tag{2.1}$$

Then, the transition dynamic of $\xi(t) = \{S_1(t), E_1(t), H_1(t), R_1(t), \ldots, S_n(t), E_n(t), H_n(t), R_n(t)\} \in \{0, 1, 2, ..., N\}^{4n}$ over $t \in [0, T]$ is described below:

- Transmission in the k-th region: A case from $E_k(t)$ chooses an individual from $N_k(t)$ randomly at Poisson rate $\lambda_k(t)$, and the individual chosen is infected if it is of state $S_k(t)$. Note that this is different from traditional SEIR models, since we assume that for the COVID-19 case, the pre-symptomatic patients from $E_k(t)$ can be contagious.
- Hospitalization in the k-th region: Each individual in $E_k(t)$ will be hospitalized with Poisson rate δ.

- Recovery or death in the k-th region: Each individual in $H_k(t)$ will transfer into $R_k(t)$ with Poisson rate γ_k. The recovery rate γ_k owns spatial heterogeneity due to the uneven distribution of medical resources.

- Transportation between regions: At the end of the t-th day, all the individuals in region i except the ones in $H_i(t)$ have the same probability to travel from the i region to the j region, and the total traveling volume from the i region to the j region is $w_{i,j}^t$. Thus, if we denote $\{\xi_{ij}^{[S,t]}, \xi_{ij}^{[E,t]}, \xi_{ij}^{[R,t]}\}$ as the number of people transported from $\{S_i(t), E_i(t), R_i(t)\}$ to $\{S_j(t), E_j(t), R_j(t)\}$ at the end of the t-th day, then $\{\xi_{ij}^{[S,t]}, \xi_{ij}^{[E,t]}, \xi_{ij}^{[R,t]}\}$ follows a multinomial distribution. Specifically,

$$P(\{\xi_{ij}^{[S,t]}, \xi_{ij}^{[E,t]}, \xi_{ij}^{[R,t]}\}) = \frac{(w_{i,j}^t)!}{\left(\xi_{ij}^{[S,t]}\right)! \left(\xi_{ij}^{[E,t]}\right)! \left(\xi_{ij}^{[R,t]}\right)!} \left(\frac{S_i(t)}{(N_a)_i(t)}\right)^{\xi_{ij}^{[S,t]}} \left(\frac{E_i(t)}{(N_a)_i(t)}\right)^{\xi_{ij}^{[E,t]}} \left(\frac{R_i(t)}{(N_a)_i(t)}\right)^{\xi_{ij}^{[R,t]}}, \tag{2.2}$$

  with $\xi_{ij}^{[S,t]} + \xi_{ij}^{[E,t]} + \xi_{ij}^{[R,t]} = w_{i,j}^t$.

For clarification, we list all the notations and the parameters in Table 1 below.

Now we consider the mean-field differential equation system in (2.3) ([25, 26]), which is a deterministic version of the stochastic model above:

$$\begin{cases} \dfrac{d\tilde{S}_k(t)}{dt} = -\lambda_k \dfrac{\tilde{S}_k \tilde{E}_k(t)}{\tilde{N}_k(t)} + \sum\limits_{i \neq k} \left[w_{i,k}^t \dfrac{\tilde{S}_i(t)}{(\tilde{N}_a)_i(t)} - w_{k,i}^t \dfrac{\tilde{S}_k(t)}{(\tilde{N}_a)_k(t)} \right] \\ \dfrac{d\tilde{E}_k(t)}{dt} = \lambda_k \dfrac{\tilde{S}_k(t)\tilde{E}_k(t)}{\tilde{N}_k(t)} - \delta \tilde{E}_k(t) + \sum\limits_{i \neq k} \left[w_{i,k}^t \dfrac{\tilde{E}_i(t)}{(\tilde{N}_a)_i(t)} - w_{k,i}^t \dfrac{\tilde{E}_k(t)}{(\tilde{N}_a)_k(t)} \right] \\ \dfrac{d\tilde{H}_k(t)}{dt} = \delta \tilde{E}_k(t) - \gamma_k \tilde{H}_k(t) \\ \dfrac{d\tilde{R}_k(t)}{dt} = \gamma_k \tilde{H}_k(t) + \sum\limits_{i \neq k} \left[w_{i,k}^t \dfrac{\tilde{R}_i(t)}{(\tilde{N}_a)_i(t)} - w_{k,i}^t \dfrac{\tilde{R}_k(t)}{(\tilde{N}_a)_k(t)} \right] \\ \dfrac{d(C_a)_k(t)}{dt} = \delta \tilde{E}_k(t) \\ \dfrac{d(R_a)_k(t)}{dt} = \gamma_k \tilde{H}_k(t) \end{cases}, \tag{2.3}$$

where $\tilde{S}_k(t)$ is the deterministic counterpart of $S_k(t)$ and so on, and $(C_a)_k(t)$ and $(R_a)_k(t)$ are the accumulated confirmed cases and the accumulated removed cases determined by the ordinary differential equation (ODE) system (2.3) in the k-th region at time t, respectively.

Among the compartments in (2.3), $\{(C_a)_k(i)\}_{i=1}^T$ and $\{(R_a)_k(i)\}_{i=1}^T$ can actually be observed, whose observed counterparts are denoted as $\{\widehat{(C_a)}_k(i)\}_{i=1}^T$ and $\{\widehat{(R_a)}_k(i)\}_{i=1}^T$, respectively.

For inference of parameters, we further denote $(\Delta C_a)_k(i) = (C_a)_k(i) - (C_a)_k(i-1)$ as the newly confirmed cases on the i-th day ($k = 1, \ldots, n, i = 2, \ldots, T$), and $\widehat{(\Delta C_a)}_k(i)$ as the observable counterparts of $(\Delta C_a)_k(i)$. The same convention holds for the definitions of $(\Delta R_a)_k(i)$ and $\widehat{(\Delta R_a)}_k(i)$.

For the remaining latent variables, $R_k(0)$ is assumed to be 0, while $E_k(0)$ and $H_k(0)$ are left to be inferred for each $k = 1, \ldots, n$.

Note that in contrast to the work based on GLEaM [11, 3, 4, 5, 14], we allow parameters to possess both spatial and temporal heterogeneity. Moreover, the transportation between regions is also considered in the model when the traveling matrix W_t is available, which is similar to GLEaM.

We finally emphasize that the model and the inference algorithm described below can be applied when the newly confirmed and recovered cases, the transportation matrix in each day, and the timeline of significant policy responses to the pandemic are available.

Table 1: List of notations and parameters

Notation	Description	Notes
$S_k(t)$	Susceptible individuals in the k-th region at time $t \in [0, T]$	
$E_k(t)$	Exposed individuals in the k-th region at time $t \in [0, T]$	$\{E_k(0)\}_{k=1}^n$ need to be inferred.
$H_k(t)$	Hospitalized individuals in the k-th region at time $t \in [0, T]$	$\{H_k(0)\}_{k=1}^n$ need to be inferred.
$R_k(t)$	Removed individuals in the k-th region at time $t \in [0, T]$	$R_k(0) = 0, \forall k$.
$N_k(t)$	Total individuals in the k-th region at time $t \in [0, T]$	$\sum_{k=1}^n N_k(t)$ is constant over time.
W_t	The traveling volume matrix on the t-th day, $t \in \{1, \ldots, T\}$	
$\tilde{S}_k(t)$	The deterministic counterpart of $S_k(t)$	Other compartments follow the same convention of notations.
$(C_a)_k(t)$	The accumulated confirmed cases determined by (2.3)	
$(R_a)_k(t)$	The accumulated recovered cases determined by (2.3)	
$(\widehat{C_a})_k(i)$	Observable counterpart of $(C_a)_k(i)$ on the i-th day	
$(\widehat{R_a})_k(i)$	Observable counterpart of $(R_a)_k(i)$ on the i-th day	
$(\Delta C_a)_k(i)$	The newly confirmed cases determined by (2.3) on the i-th day	
$(\widehat{\Delta C_a})_k(i)$	Observable counterpart of $(\Delta C_a)_k(i)$ on the i-th day	
λ_k	Infection rate in the k-th region	$\{\lambda_k\}_{k=1}^n$ need to be inferred, and are allowed to be time-varying.
δ	Inverse of average incubation period	δ is prefixed as 0.14.
γ_k	Inverse of average recovery / death time	$\{\gamma_k\}_{k=1}^n$ are inferred first by the newly recovered cases and the total hospitalized cases.
Θ	Set of parameters to be estimated	
W	The matrix characterizing the proximity between regions	
$D_m, 1 \leq m \leq d$	Group of regions	
μ_0	Parameter characterizing the correlation between groups	
$\mu_m, 1 \leq m \leq d$	Parameter characterizing the correlation within group m	
A	Affinity matrix constructed by W and $\{\mu_m\}_{m=0}^d$	
σ	Parameter in the prior (2.7)	

2.2 Estimation of Model Parameters

A two-step procedure of estimating parameters is described below.

First, note that δ, the inverse of the average time for a person from being exposed to hospitalized, is prefixed to be $\delta = 0.14$ for all regions and all time. According to [27], the mean duration of incubation period is 5.2 days. Furthermore, we assume that the average time for an individual from showing symptoms to being hospitalized is 2 days for simplicity. Thus, the mean duration for an individual from being exposed to being hospitalized is 7.2 days, whose inverse value is approximately 0.14.

Thus, the model parameters that need to be estimated are $\Theta = \{E_k(0), H_k(0), \lambda_k, \gamma_k\}_{k=1}^n$.

A two-step analysis is adapted for the estimation. We first make inference for $\{\gamma_k\}_{k=1}^n$ utilizing the newly recovered cases and the accumulated hospitalized cases. After the estimation of $\{\gamma_k\}_{k=1}^n$, $\widetilde{\Theta} := \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$ are then estimated by maximizing the posterior distribution, where we introduce a prior distribution combining the information of correlation between regions. In addition, the marginal posterior distributions of $\widetilde{\Theta}$ can also be obtained by running MCMC sampling starting from the optimizer $\widetilde{\Theta}^*$ of the posterior distribution. Then we complete the inference of $\Theta = \{\{\gamma_k\}_{k=1}^n, \widetilde{\Theta}\}$.

Mathematically, given parameters Θ, we can run the ODE system (2.3) on $[0, T]$ and denote $x_i = (\{(\Delta C_a)_k(i)\}_{k=1}^n, \{(C_a)_k(i)\}_{k=1}^n, \{(R_a)_k(i)\}_{k=1}^n, \{(\Delta R_a)_k(i)\}_{k=1}^n)$ as the values determined by (2.3) on the i-th day, and $y_i = (\{(\widehat{\Delta C_a})_k(i)\}_{k=1}^n, \{(\widehat{C_a})_k(i)\}_{k=1}^n, \{(\widehat{\Delta R_a})_k(i)\}_{k=1}^n, \{(\widehat{R_a})_k(i)\}_{k=1}^n)$ as the observed counterpart of x_i.

Step 1: Estimate $\{\gamma_k\}_{k=1}^n$.

We first estimate $\{\gamma_k\}_{k=1}^n$ by maximizing $\prod_{i=1}^T P\left(\left\{(\widehat{\Delta R_a})_k(i)\right\}_{k=1}^n \middle| \{\gamma_k\}_{k=1}^n, y_{i-1}\right)$ over γ_k for each k. By assuming that the newly recovered cases in one day follow a Poisson distribution whose mean equals to the product of γ_k and the accumulated hospitalized cases (which is the difference between the accumulated confirmed cases and the accumulated recovered cases, and thus is observable) the day before, we have that

$$
\begin{aligned}
&\prod_{i=1}^T P\left(\left\{(\widehat{\Delta R_a})_k(i)\right\}_{k=1}^n \middle| \{\gamma_k\}_{k=1}^n, y_{i-1}\right) \\
=&\prod_{i=1}^T \prod_{k=1}^n \mathrm{Pois}\left((\widehat{\Delta R_a})_k(i) \middle| \gamma_k\left((\widehat{C_a})_k(i-1) - (\widehat{R_a})_k(i-1)\right)\right) \\
=&\prod_{k=1}^n \prod_{i=1}^T \frac{\left(\gamma_k\left((\widehat{C_a})_k(i-1) - (\widehat{R_a})_k(i-1)\right)\right)^{(\widehat{\Delta R_a})_k(i)}}{\left((\widehat{\Delta R_a})_k(i)\right)!} \exp\left(-\gamma_k\left((\widehat{C_a})_k(i-1) - (\widehat{R_a})_k(i-1)\right)\right).
\end{aligned}
$$

Then, we estimate $\gamma_k^* = \arg\max_{\gamma_k} \prod_{i=1}^T \mathrm{Pois}\left((\widehat{\Delta R_a})_k(i) \middle| ((\widehat{C_a})_k(i-1) - (\widehat{R_a})_k(i-1))\gamma_k\right)$ for each k separately.

Step 2: Estimate $\widetilde{\Theta} = \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$ and the marginal posterior distributions of $\widetilde{\Theta}$.

Next, we estimate the remaining parameters $\widetilde{\Theta} = \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$ as follows.

First, note that λ_k is the transmission rate, which reflects the infectivity of the virus, the prevention and control measures, and the rates of close contacts in the k-th region. In this report, λ_k is allowed to be time-varying. Specifically, the total T days are split into two periods with time thresholds T_k chosen for each $k = 1, \ldots, n$, then $\lambda_k = \lambda_{k,1}$ if $t < T_k$ and $\lambda_k = \lambda_{k,2}$ otherwise.

Notice that the ODE system (2.3) is the mean-field version of our stochastic model, and $\{x_i\}_{i=1}^T$ (especially $\{(\Delta C_a)_k(i)\}_{k,i}$) are determined by the parameters, thus by the Markov property, $\{(\widehat{\Delta C_a})_k(i)\}$

are all independent for $k = 1, \ldots, n$, $i = 1, \ldots, T$ conditioned on the parameters $\widetilde{\Theta}$. Furthermore we suppose that $(\widehat{\Delta C_a})_k(i) \sim \text{Pois}((\Delta C_a)_k(i))$. Thus, the likelihood of $\widetilde{\Theta}$ can be written as

$$P\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i} | \widetilde{\Theta}\right) = \prod_{i=1}^{T} P\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k=1}^{n} \middle| \widetilde{\Theta}\right) = \prod_{i=1}^{T}\prod_{k=1}^{n} \frac{((\Delta C_a)_k(i))^{(\widehat{\Delta C_a})_k(i)}}{((\widehat{\Delta C_a})_k(i))!} e^{-(\Delta C_a)_k(i)}. \quad (2.4)$$

Now, we denote the posterior distribution of $\widetilde{\Theta}$ as $\pi(\widetilde{\Theta})$. Then by Baysian formula,

$$\pi(\widetilde{\Theta}) = p\left(\widetilde{\Theta} \middle| \left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i}\right) = \frac{1}{Z} p\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i} \middle| \widetilde{\Theta}\right) p(\widetilde{\Theta}), \quad (2.5)$$

where $p(\widetilde{\Theta})$ is the prior distribution of $\widetilde{\Theta}$ to be determined and $Z = p\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i}\right)$ is a constant irrelevant to $\widetilde{\Theta}$. We further denote $V(\widetilde{\Theta}) = -\log\left(p\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i} \middle| \widetilde{\Theta}\right)\right) - \log(p(\widetilde{\Theta}))$, then

$$\pi(\widetilde{\Theta}) = \frac{1}{Z}\exp(-V(\widetilde{\Theta})). \quad (2.6)$$

To fit the realistic evolution of the epidemic more precisely, $\widetilde{\Theta}$ is estimated as $\widetilde{\Theta}^* = \arg\max \pi(\widetilde{\Theta}) = \arg\min V(\widetilde{\Theta})$ with reasonable prior distribution $p(\widetilde{\Theta})$. Then, MCMC sampling scheme starting from $\widetilde{\Theta}^*$ is applied to get the posterior distribution for $\widetilde{\Theta}$. This process might possess higher computational efficiency than choosing the initial point for MCMC randomly or empirically.

Choosing prior distribution

The remaining problem is to choose the prior distribution $p(\widetilde{\Theta})$. Presuming that the transmission rates in the regions owning more similarities are closer, $p(\widetilde{\Theta})$ is designed to combine the information of correlations between regions. In particular, given a matrix A which characterizes the pairwise similarities between the regions, and if we denote $\lambda^{(\mathbf{1})} = (\lambda_{1,1}, \ldots, \lambda_{n,1})^T, \lambda^{(\mathbf{2})} = (\lambda_{1,2}, \ldots, \lambda_{n,2})^T \in \mathbb{R}^n$,

$$p(\widetilde{\Theta}) = \frac{1}{C_{\sigma,A}} \exp(-{\lambda^{(\mathbf{1})}}^T (D-A)\lambda^{(\mathbf{1})} - \sigma\|\lambda^{(\mathbf{1})}\|_2^2 - {\lambda^{(\mathbf{2})}}^T (D-A)\lambda^{(\mathbf{2})} - \sigma\|\lambda^{(\mathbf{2})}\|_2^2), \quad (2.7)$$

where $D = \text{diag}\{d_1, \ldots, d_n\}$ is the degree matrix of A with $d_i = \sum_{j=1}^{i} A_{ij}$, $C_{\sigma,A} = \int_{\mathbb{R}^{2n}} \exp(-{\lambda^{(\mathbf{1})}}^T (D - A)\lambda^{(\mathbf{1})} - \sigma\|\lambda^{(\mathbf{1})}\|_2^2 - {\lambda^{(\mathbf{2})}}^T (D-A)\lambda^{(\mathbf{2})} - \sigma\|\lambda^{(\mathbf{2})}\|_2^2) d\lambda^{(\mathbf{1})} d\lambda^{(\mathbf{2})}$ is a constant depending on σ and A. Here, a small σ is chosen for $p(\widetilde{\Theta})$ to be a probability measure without imposing much restriction on $\lambda^{(1)}$ or $\lambda^{(2)}$. Then,

$$\begin{aligned} V(\widetilde{\Theta}) = &-\log\left(p\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i} \middle| \widetilde{\Theta}\right)\right) + {\lambda^{(\mathbf{1})}}^T (D-A)\lambda^{(\mathbf{1})} + \sigma\|\lambda^{(\mathbf{1})}\|_2^2 + {\lambda^{(\mathbf{2})}}^T (D-A)\lambda^{(\mathbf{2})} \\ &+ \sigma\|\lambda^{(\mathbf{2})}\|_2^2 + \log C_{\sigma,A} \\ = &-\log\left(p\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i} \middle| \widetilde{\Theta}\right)\right) + \sum_{i<j} a_{ij}(\lambda_{i,1} - \lambda_{j,1})^2 + \sigma\|\lambda^{(\mathbf{1})}\|_2^2 + \sum_{i<j} a_{ij}(\lambda_{i,2} - \lambda_{j,2})^2 \\ &+ \sigma\|\lambda^{(\mathbf{2})}\|_2^2 + \log C_{\sigma,A}. \end{aligned} \quad (2.8)$$

Recalling that

$$\begin{aligned} \widetilde{\Theta}^* &= \arg\min V(\widetilde{\Theta}) \\ &= \arg\min(-\log\left(p\left(\left\{(\widehat{\Delta C_a})_k(i)\right\}_{k,i} \middle| \widetilde{\Theta}\right)\right) + \sum_{i<j} a_{ij}(\lambda_{i,1} - \lambda_{j,1})^2 + \sigma\|\lambda^{(\mathbf{1})}\|_2^2 \\ &\quad + \sum_{i<j} a_{ij}(\lambda_{i,2} - \lambda_{j,2})^2 + \sigma\|\lambda^{(\mathbf{2})}\|_2^2), \end{aligned} \quad (2.9)$$

it can be seen that by choosing the prior distribution as in (2.7), a l_2 regularization term $\frac{1}{2}\sum_{i,j} a_{ij}(\lambda_{i,1} - \lambda_{j,1})^2 + \frac{1}{2}\sum_{i,j} a_{ij}(\lambda_{i,2} - \lambda_{j,2})^2$ is imposed for better generalization.

A is constructed from another affinity matrix W by further addressing the correlation between regions with more similarities.

For simulated data and real-world data in China, in which cases the transportation data are available, we construct W from $\{W_i\}_{i=1}^T$. Specifically, $W := \frac{1}{2}((\frac{1}{T}\sum_{i=1}^T W_i) + (\frac{1}{T}\sum_{i=1}^T W_i)^T)$. Nevertheless, for real-world data in Europe, where we are not aware of traveling data publicly available that are sufficient for the proposed model, W is just the all 1 adjacency matrix multiplied by a scalar. We remark that the affinity matrix W may also be obtained by ways other than using the transportation data, as long as it reflects the similarities between districts.

Once W is obtained, the next step of attaining A is to divide the n regions into d groups ($D_1, ..., D_d$, $\cup_{m=1}^d D_m = \{1, \ldots, n\}$, and $\forall i \neq j$, $D_i \cap D_j = \emptyset$) so that the regions in the same groups have more similarities. Then a_{ij} is constructed as follows. For any $i, j \in \{1, \ldots, n\}$, if $i, j \in D_m$ for some $m \in \{1, \ldots, d\}$, $a_{ij} = \mu_m w_{ij}$ for some $\mu_m > 0$ relatively large; otherwise, if $i, j \notin D_m$ for any $m \in \{1, \ldots, d\}$, $a_{ij} = \mu_0 w_{ij}$ for some $\mu_0 > 0$ relatively small. By constructing A in the aforementioned way, correlations for regions in the same groups are further strengthened.

Finally, after choosing $p(\widetilde{\Theta})$ by determining A and σ, the optimization process $\widetilde{\Theta}^* = \arg\min V(\widetilde{\Theta})$ is accomplished through Matlab function `fmincon`. In addition, the posterior distribution for $\widetilde{\Theta}$ could be obtained by classical MCMC sampling scheme starting from $\widetilde{\Theta}^*$ as mentioned above.

We finally summarize the procedure described in Section 2.2 in Algorithm 1 below.

Algorithm 1 GLEaM model with Graph Laplacian Prior Distribution

Input: Data of total confirmed cases and total recovered and dead cases $\{(\widehat{C_a})_k(i)\}_{k,i}, \{(\widehat{R_a})_k(i)\}_{k,i}$, traveling volume matrices $\{W_t\}_{t=1}^T$, graph Laplacian prior parameters $\mu_0, \mu_1, \ldots, \mu_d$, σ and affinity matrix W, division of n provinces $D_1, \ldots, D_d$, number of MCMC iterations N_{ite}

Output: Inferred parameters Θ^*, estimated marginal posterior distributions of Θ, predicted deterministic trajectory $\hat{y}_{1:T}$

Compute newly confirmed cases and newly recovered and dead cases, $(\widehat{\Delta C_a})_k(i) \leftarrow (\widehat{C_a})_k(i) - (\widehat{C_a})_k(i-1)$, $(\widehat{\Delta R_a})_k(i) \leftarrow (\widehat{R_a})_k(i) - (\widehat{R_a})_k(i-1)$

Compute $\gamma_k^* \leftarrow \arg\max_{\gamma_k} \prod_{i=1}^T \text{Pois}\left((\widehat{\Delta R_a})_k(i) \middle| ((\widehat{C_a})_k(i-1) - (\widehat{R_a})_k(i-1))\gamma_k\right)$

Construct another affinity matrix A by

for $i, j = 1, \ldots, n$ **do**

  if $i, j \in D_m$ for some $m \in \{1, \ldots, d\}$ **then**

    $a_{ij} \leftarrow \mu_m w_{ij}$

  else

    $a_{ij} \leftarrow \mu_0 w_{ij}$

  end if

end for

Compute the degree matrix D, $D_{ii} \leftarrow \sum_j a_{ij}$

Compute the inferred parameters $\widetilde{\Theta}^* \leftarrow \arg\min_{\widetilde{\Theta}} V(\widetilde{\Theta}) = \arg\min_{\widetilde{\Theta}} \left(-\log\left(p\left(\{(\widehat{\Delta C_a})_k(i)\}_{k,i} \middle| \widetilde{\Theta}\right)\right) + \lambda^T(D-A)\lambda + \sigma\|\lambda\|_2^2\right)$. $V(\widetilde{\Theta})$ should be changed accordingly if λ is allowed to be time-varying

$\Theta^* = \{\{\gamma_k^*\}_{k=1}^n, \widetilde{\Theta}^*\}$

14: Compute the predicted deterministic trajectory $\hat{y}_{1:T}$ by applying explicit Euler's method to (2.3) with Θ^*

15: Run MCMC staring from $\widetilde{\Theta}^*$ for N_{ite} iterations

16: Compute the estimated marginal posterior distribution of $\theta \in \widetilde{\Theta}$ by $\hat{p}(\theta|\{(\widehat{\Delta C_a})_k(i)\}_{k=1\ldots,n,i=1,\ldots,T}) \leftarrow \sum_{l=1}^{n_{bin}} \frac{c_l}{N} \mathbf{1}_{\theta \in H_l}$, where $\{H_l\}_{k=1}^{n_{bin}}$ are the bins and $\{c_l\}_{k=1}^{n_{bin}}$ are the numbers of elements in the bins in N_{ite} iterations

2.3 Prediction of the Epidemic Trajectories with the Estimated Parameters

Once the estimated parameters $\Theta^* = \{\{\gamma_k^*\}_{k=1}^n, \widetilde{\Theta}^*\}$ are obtained from the optimizations $\gamma_k^* = \arg\max_{\gamma_k} \prod_{i=1}^T \text{Pois}\left((\widehat{\Delta R_a})_k(i) \middle| ((\widehat{C_a})_k(i-1) - (\widehat{R_a})_k(i-1))\gamma_k\right)$ and $\widetilde{\Theta}^* = \arg\min V(\widetilde{\Theta})$ as described in Section 2.2, the trajectories of newly confirmed cases could be simulated according to the stochastic dynamic process with Θ^*. Furthermore, trajectories could also be sampled from the posterior distribution of Θ instead of using Θ^* alone, which also takes the randomness from $\widetilde{\Theta}$ into account. Particularly, this could be achieved by sampling $\widetilde{\Theta}$ from MCMC and then simulating trajectories with the sampled $\{\{\gamma_k^*\}_{k=1}^n, \widetilde{\Theta}\}$. Additionally, deterministic trajectories determined by (2.3) could also be computed by explicit Euler's method as in Algorithm 1.

3 Experimental Results for simulated data

Two specific cases are considered for simulated data. The first one considers only two provinces with transportation between each other. In contrast, the other case considers four provinces separated into two groups (the provinces in the same group are assumed to have more similarities) and with traffic between each pair of the provinces.

For each case, we first compare the predicted trajectories of our dynamic model with four baseline models detailed in Section 3.1.2. Comparison between the baseline models shows that heterogeneity of transmission parameters and transportation in the model improves the prediction. Then, the comparison between the baseline models and the proposed model, which further involves the correlation between regions besides heterogeneity and transportation, indicates that the graph Laplacian regularization improves generalization performance. Furthermore, the estimation of transmission parameters using different models shows the benefits of involving heterogeneity of transmission parameters. This also shows that the proposed model can estimate the transmission parameters, validating the capability of the model proposed in this paper in predicting the trajectories. Finally, we show quantitative evaluations of the proposed model and the baseline models by comparing training errors and testing errors defined in Appendix A, which further verifies our findings above.

We remark that the results below are for 100 replicas. In each replica, two random trajectories are sampled according to the stochastic model with prefixed parameters, part of which are treated as the ground truth training and testing trajectory, respectively (see more details in Appendix A). For each model, we first fit the parameters using the training trajectory and then predict the testing trajectory using the estimated parameters. The comparison of trajectory prediction is from one typical realization, for which we compare the ground truth training and testing trajectories with the fitted training and predicted testing trajectories for all the models. Additionally, parameter estimation and quantitative evaluations are compared with mean and standard deviation over all 100 replicas. The detailed computations of training and testing errors for simulated data can be found in Appendix A.

3.1 Two Provinces Case

3.1.1 Data Description

In this case, $n = 2$. The data span over a period of $T = 30$ days, and the threshold T_{th} separating training and testing data is set to be 10 (more detailed can be seen in Appendix A). The other prefixed parameters are as follows:

- For $k \in \{1, 2\}$, $N_k(0) = 10^6$, $E_k(0) = 30$, $H_k(0) = 10$.
- For $i \in \{1, \ldots, T\}$, $W_{12}^i = W_{21}^i = w_{12} = 5 \times 10^3$.
- $\lambda_1 = 0.4$, $\lambda_2 = 0.35$, $\delta = \gamma_1 = \gamma_2 = 0.14$. $\{\lambda_k\}$ are not time-varying.

3.1.2 Models compared

We first specify the models to be compared below. The last one is the proposed model, and the first four models serve as baselines with different settings.

1. Models with uniform prior distribution, without heterogeneity or migration.
2. Models with uniform prior distribution, without heterogeneity but with migration.
3. Models with uniform prior distribution, with heterogeneity but without migration.
4. Models with uniform prior distribution, with both heterogeneity and migration.
5. Models with prior distribution based on graph Laplacian, with both heterogeneity and migration.

First, the models with uniform prior distributions themselves (Models 1-4) are compared according to whether two key assumptions exist in the model:

(1) Whether the transmission rates $\{\lambda_k\}$ are allowed to vary over regions.

(2) Whether there exists transportation between regions.

Then, the model with prior distribution based on graph Laplacian (Model 5) is compared with those using uniform distributions as prior distributions (Models 1-4). The former one utilizes the correlation between subpopulations by adding a l_2 regularization term for the model. In contrast, only lower and upper bounds are imposed on parameters without other prior information being used in the latter ones.

Note that for Model 5, that utilizes graph Laplacian based prior distribution as defined in (2.7), the parameter σ is taken to be 10. Moreover, since there are only two provinces in the current experiment, the intra-group parameter $\{\mu_m, m = 1, 2\}$ does not appear in the final penalty, and only the inter-group parameter determines the graph Laplacian penalty, which is denoted as μ. Specifically, by (2.9), $\widetilde{\Theta} = \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$ are estimated by the optimization problem in (3.1):

$$\widetilde{\Theta}^* = \arg\min(-\log p(y_{1:T_{th}}|\widetilde{\Theta}) + \mu w_{12}(\lambda_1 - \lambda_2)^2 + \sigma\|\lambda\|_2^2). \tag{3.1}$$

We remark that in Models 1-4, the estimation of $\Theta = \{\{\gamma_k\}_{k=1}^n, \widetilde{\Theta}\}$ still follows a similar two-step procedure as in Model 5, and the first step of obtaining $\gamma_k^* = \arg\max_{\gamma_k} \prod_{i=1}^T \mathrm{Pois}\big((\widehat{\Delta R_a})_k(i)\big|((\widehat{C_a})_k(i-1) - (\widehat{R_a})_k(i-1))\gamma_k\big)$ remains formally the same. The difference lies in the optimization object of $\widetilde{\Theta}$. First, the l_2 regularization term becomes prior knowledge of the parameters' upper and lower bounds. Second, for models without heterogeneity of parameters, $\{\lambda_k\}_{k=1}^n$ are forced to be the same in (2.3). For models

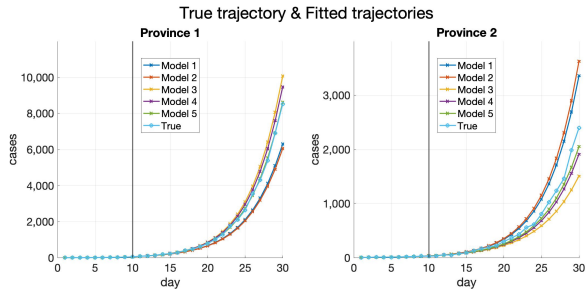

Figure 1: True and fitted trajectories of Models 1-5 in the two provinces. The vertical lines show the threshold $T_{th} = 10$.

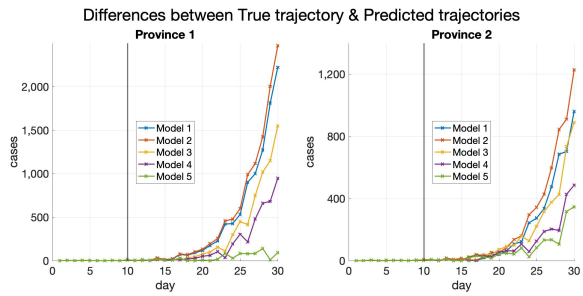

Figure 2: Differences of true and fitted trajectories of Models 1-5 in the two provinces. The vertical lines show the threshold $T_{th} = 10$.

without transportation between regions, terms involving W_t disappear in (2.3). Additionally, for the other data sets considered in the following sections, the parameter estimation methods for the baseline models are similar and thus will not be repeated.

Finally, note that the model in [11] is similar to Models 1 and 2, since they all assume a spatially homogeneous transmission parameter. However, [11] assumed that the epidemic was seeded from one seed region while Models 1 and 2 do not make such assumption. Moreover, [11] focused more on the spread of the epidemic from the seed region at the early stage of the pandemic, and only the introduction dates in the other regions were utilized for the estimation of transmission parameters. In comparison, the estimation of transmission rate in Models 1 and 2 exploits the data in all regions in the whole process.

3.1.3 Results of Trajectory Comparison

The true trajectories from a typical realization and the predicted trajectories of Models 1-5 are plotted in Figure 1, and the differences of the true trajectories and the fitted trajectories are plotted in Figure 2 for better comparison of the five models.

As can be seen in Figure 1 and 2, Model 4 which allows both heterogeneity and transportation fits the training data better and also predicts the testing data with relatively smaller absolute errors compared with Models 1-3. In addition, Model 5, which further combines the correlation between provinces into prior distribution, has better performance of generalization than Model 4.

3.1.4 Results of Parameter Estimation

The mean and standard deviation of $\{\lambda_i\}$ estimated by the five models are reported in Table 2. It can be seen from Table 2 that models allowing heterogeneity estimate parameters more accurately, and Model 5 that integrates the correlation leads to slightly better estimation for λ_2.

Model	M	H	GL prior	λ_1	λ_2
1	×	×	×	0.3798 ± 0.0304	
2	✓	×	×	0.3798 ± 0.0304	
3	×	✓	×	0.3922 ± 0.0426	0.3415 ± 0.0412
4	✓	✓	×	0.3933 ± 0.0431	0.3396 ± 0.0418
5	✓	✓	✓	0.3907 ± 0.0390	0.3500 ± 0.0339

Table 2: Estimated λ_i of five models. Recall that the ground truth is $\lambda_1 = 0.4$, $\lambda_2 = 0.35$. The first three columns indicate whether the model permits **migration** between provinces, the **heterogeneity** of parameters and the **Graph Laplacian prior**, respectively. $\{\lambda_i\}$ are inferred for each of the 100 replicas, then the mean and standard deviation are presented in the table above. In addition, $\mu = 10^{-2}$ in Model 5 for inference of parameters as defined in (3.1).

3.1.5 Further Model Evaluation

The training and testing errors computed according to Appendix A are listed in Table 3 below. From Table 3, it can be seen that even without the prior based on graph Laplacian, heterogeneity of parameters and transportation between two provinces help reduce the testing errors. The concurrence of both the two assumptions makes the training and testing errors further decrease.

Model	M	H	GL prior	TrainRelErr$^{(1)}$	TestRelErr$^{(1)}$	TrainRelMSE$^{(1)}$	TestRelMSE$^{(1)}$
1	×	×	×	0.3895 ± 0.0887	1.0289 ± 0.6646	0.5462 ± 0.1318	1.0644 ± 0.6946
2	✓	×	×	0.3938 ± 0.0899	1.0838 ± 0.6828	0.5507 ± 0.1327	1.1214 ± 0.7150
3	×	✓	×	0.3645 ± 0.0859	0.9694 ± 0.5877	0.5142 ± 0.1252	0.9953 ± 0.6043
4	✓	✓	×	0.3640 ± 0.0861	0.9508 ± 0.5811	0.5136 ± 0.1250	0.9756 ± 0.5987
5	✓	✓	✓	0.3641 ± 0.0863	0.8782 ± 0.5890	0.5153 ± 0.1262	0.9030 ± 0.6082

Table 3: Training and testing error with standard deviation for simulated data (two provinces) of five models. The first three columns indicate whether the model permits **migration** between provinces, the **heterogeneity** of parameters and the **Graph Laplacian prior**, respectively. TrainRelErr$^{(1)}$, TestRelErr$^{(1)}$, TrainRelMSE$^{(1)}$ and TestRelMSE$^{(1)}$ are computed for each of the 100 replicas, then the mean and standard deviation are presented in the table above. In addition, $\mu = 10^{-2}$ in Model 5 for inference of parameters as defined in (3.1).

The plots of the mean of TestRelErr$^{(1)}$ and TestRelErr$^{(2)}$ against varying μ are shown in Figure 3. First, note that TestRelErr$^{(1)}$ of Model 4 is nearly the same as that of Model 5 with small μ, since Model 4 is the same as Model 5 with $\mu = 0$. Then, it can be observed from the trend of TestRelErr$^{(1)}$ that as μ decreases, the mean of TestRelErr$^{(1)}$ first decreases and then slightly increases. This indicates that as μ decreases, Model 5 first underfits when μ is large since it forces $\{\lambda_k\}$ to be close to each other, and

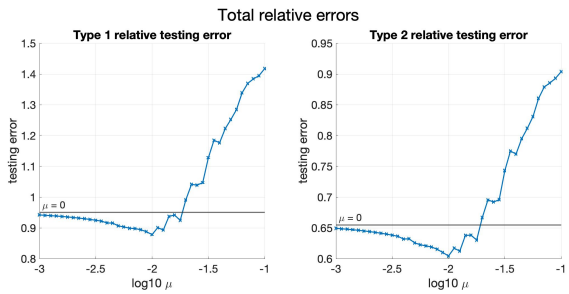

Figure 3: Mean of $\text{TestRelErr}^{(1)}$ and $\text{TestRelErr}^{(2)}$ against μ for simulated data (two provinces). The horizontal green lines show the values of the mean of the testing errors when $\mu = 0$.

then overfits when μ is small since nearly no restriction is imposed on $\{\lambda_k\}$. $\text{TestRelErr}^{(1)}$ achieves its minimum when μ is close to 10^{-2}.

Figure 17 in Appendix C shows relative mean square testing errors against varying μ, in which the same trends described above can be seen.

However, the testing errors are close when μ is relatively small, and thus the benefit of introducing graph Laplacian based prior is less significant in this case. Such a trend is more apparent in the four provinces case in the following subsection.

3.2 Four provinces case

3.2.1 Data Description

In this simulated study, we let $n = 4$, $T = 30$, $T_{th} = 10$. The other prefixed parameters are listed below:

- For $k \in \{1, 2, 3, 4\}$, $N_k(0) = 10^6$, $E_k(0) = 30$, $H_k(0) = 10$.
- For $i \in \{1, \ldots, T\}$, $k, l \in \{1, 2, 3, 4\}$ and $k \neq l$, $W^i_{kl} = 5 \times 10^3$.
- $\lambda_1 = 0.5$, $\lambda_2 = 0.47$, $\lambda_3 = 0.4$, $\lambda_4 = 0.37$, $\delta = \gamma_1 = \cdots = \gamma_4 = 0.14$. $\{\lambda_k\}$ are not time-varying.

The four provinces are divided into two groups, with the first group consisting of province 1 and 2 and the second group consisting of province 3 and 4. The similarities within groups are reflected in the settings that the values of $\{\lambda_k\}$ are closer for provinces in the same group.

3.2.2 Models compared

The same five models (Models 1-5) as described in Section 3.1.2 are compared for the four provinces case.

Notice that for Model 5, there are three parameters μ_0, μ_1, μ_2 in the graph Laplacian prior, among which μ_0 controls the inter-group correlation, μ_1 and μ_2 control the correlations within group 1 and group 2, respectively. For simplicity of model comparison, we let $\mu_0 = 0$ (no restrains on λ_i's in different groups), and $\mu_1 = \mu_2$. The parameter σ in (2.7) is still chosen to be 10. Then for Model 5, by (2.9), $\widetilde{\Theta} = \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$ are estimated by the optimization problem in (3.2):

$$\widetilde{\Theta}^* = \arg\min(-\log p(y_{1:T_{th}}|\widetilde{\Theta}) + \mu_1 w_{12}\left((\lambda_1 - \lambda_2)^2 + (\lambda_3 - \lambda_4)^2\right) + \sigma\|\lambda\|_2^2). \tag{3.2}$$

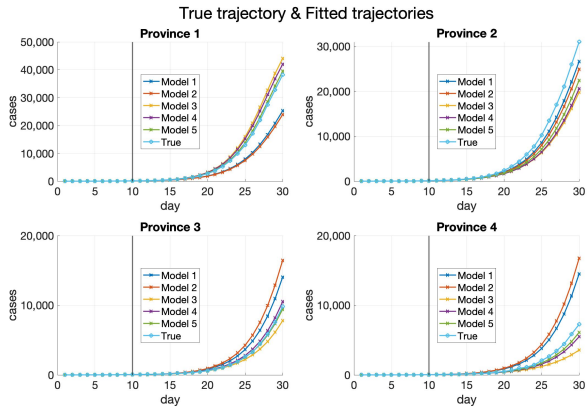

Figure 4: True and fitted trajectories of Models 1-5 in the four provinces. The vertical lines show the threshold $T_{th} = 10$.

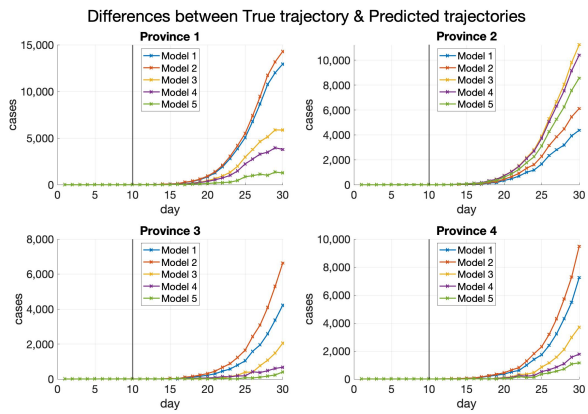

Figure 5: Differences of true and fitted trajectories of Models 1-5 in the four provinces. The vertical lines show the threshold $T_{th} = 10$.

3.2.3 Results of Trajectory Comparison

The trajectories of a typical realization are plotted in Figure 4 and the differences between the true and fitted trajectories are shown in Figure 5. As shown in these two figures, heterogeneity helps improve the prediction of testing data more than transportation, while introducing migration without heterogeneity of parameters worsens the estimation as can also be noticed from Table 5. More explanations can be found in Section 3.2.5.

Additionally, Model 5 with prior distrbution based on graph Laplacian lowers the absolute errors of predicted trajectories compared with Model 4.

3.2.4 Results of Parameter Estimation

The mean and standard deviation of $\{\lambda_i\}$ estimated by the five models for four provinces case are reported in Table 4. The same trend as in the two provinces case is observed in the current settings.

Model	M	H	GL prior	λ_1	λ_2	λ_3	λ_4
1	×	×	×	0.4466 ± 0.0190			
2	✓	×	×	0.4467 ± 0.0190			
3	×	✓	×	0.4840 ± 0.0358	0.4591 ± 0.0341	0.3918 ± 0.0383	0.3703 ± 0.0370
4	✓	✓	×	0.4887 ± 0.0365	0.4615 ± 0.0350	0.3861 ± 0.0401	0.3604 ± 0.0385
5	✓	✓	✓	0.4880 ± 0.0325	0.4731 ± 0.0301	0.3853 ± 0.0346	0.3721 ± 0.0306

Table 4: Estimated λ_i of five models. Recall that the ground truth is $\lambda_1 = 0.5$, $\lambda_2 = 0.47$, $\lambda_3 = 0.4$, $\lambda_4 = 0.37$. The first three columns indicate whether the model permits **migration** between provinces, the **heterogeneity** of parameters and the **Graph Laplacian prior**, respectively. λ_i are inferred for each of the 100 replicas, and then the mean and standard deviation are presented in the table above. In addition, $\mu = 10^{-1.55}$ in Model 5 for inference of parameters as defined in (3.2).

3.2.5 Further Model Evaluation

The training and testing errors are listed below in Table 5. Different from the two provinces case, it can be seen from Table 5 that the errors increase greatly after transportation is included while heterogeneity remains absent. A possible explanation for this might be that without heterogeneity of parameters and transportation between provinces, the estimated λ_k's are lower than the true λ_k's for group 1, which leads to that the estimated newly confirmed cases are fewer than the true ones in group 1. For the same reason, the estimated newly confirmed cases are higher than the true ones in group 2. When the transportation is considered, more confirmed cases in group 1 are transferred to group 2 than the cases transported in the opposite direction. As a result, when the transmission parameters do not have heterogeneity, migration between provinces will worsen the prediction performance compared to the case without migration.

Model	M	H	GL prior	$\text{TrainRelErr}^{(1)}$	$\text{TestRelErr}^{(1)}$	$\text{TrainRelMSE}^{(1)}$	$\text{TestRelMSE}^{(1)}$
1	×	×	×	0.6940 ± 0.1084	1.7547 ± 0.6653	0.9525 ± 0.1441	1.7751 ± 0.6751
2	✓	×	×	0.7288 ± 0.1154	2.3273 ± 0.7609	0.9935 ± 0.1524	2.3513 ± 0.7744
3	×	✓	×	0.5713 ± 0.0889	1.4852 ± 0.4092	0.8295 ± 0.1416	1.5087 ± 0.4137
4	✓	✓	×	0.5709 ± 0.0889	1.1916 ± 0.4708	0.8284 ± 0.1418	1.2076 ± 0.4757
5	✓	✓	✓	0.5745 ± 0.0895	1.0944 ± 0.565	0.8310 ± 0.1406	1.1109 ± 0.5736

Table 5: Training and testing error with standard deviation for simulated data (four provinces) of five models. As in the two provinces case, The first three columns indicate whether the model permits **migration** between provinces, the **heterogeneity** of parameters and the **Graph Laplacian prior**, respectively. $\text{TrainRelErr}^{(1)}$, $\text{TestRelErr}^{(1)}$, $\text{TrainRelMSE}^{(1)}$ and $\text{TestRelMSE}^{(1)}$ are still computed for each of the 100 replicas, then the mean and standard deviation are presented in the table above. In addition, $\mu_1 = 10^{-1.55}$ in Model 5 for inference of parameters as defined in (3.2).

In addition, Table 5 still shows that the presence of both heterogeneity and transportation helps reduce the training and testing errors by comparing the first four models. By comparing Model 4 and Model 5, it can be seen that using the graph Laplacian prior could further reduce the testing errors.

As in the previous subsection, the mean of $\text{TestRelErr}^{(1)}$ and $\text{TestRelErr}^{(2)}$ against varying μ_1 is shown below in Figure 6, and the mean of $\text{TestRelMSE}^{(1)}$ and $\text{TestRelMSE}^{(2)}$ is shown in Figure 18 in Appendix C. From Figure 6 and Figure 18, the phenomenon that Model 5 first underfits and then overfits as μ_1 decreases can be seen more recognizably compared with the two provinces case.

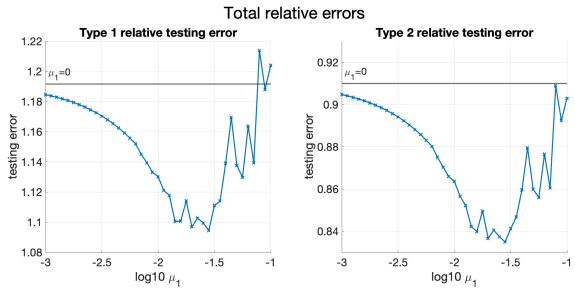

Figure 6: Mean of TestRelErr$^{(1)}$ and TestRelErr$^{(2)}$ against μ_1 for simulated data (four provinces). The horizontal green lines show the values of the mean of the testing errors when $\mu_1 = 0$.

4 Experimental Results for Real-World COVID-19 Data

Real-world COVID-19 data in China and Europe are analyzed in this paper, respectively. Similar to the simulated data cases in Section 3, we first show the predicted trajectories of the proposed model (both deterministic and stochastic ones) and the baseline models (only deterministic ones) for some selected provinces in China or countries in Europe. This shows that the proposed model can capture and further predict the trend of the true trajectories and behaves better than the baseline models, mainly due to the introduction of heterogeneity of transmission parameters and the employment of correlation between provinces or countries. After that, we show the estimation of the posterior distribution of transmission parameters in a selected province or country using the proposed model. Finally, we evaluate the models quantitatively by computing the training and testing errors as described in Appendix A, which further demonstrates the benefits of the proposed model compared to the baselines.

4.1 Data Sources

Real-world data studied in this section involve the data in China and Europe.

Data in China include the number of confirmed cases, recovered cases, and fatalities in China's major provinces and municipalities from January 21st to March 28th. These publicly available data are downloaded from websites of health commissions of the provinces. The corresponding total population in each province or municipality is from [28].

Transportation data in China are also utilized, lasting from January 10th to February 10th. The data used to reflect the number of people that move out from the cities are collected from Baidu Qianxi [29] (provided by Baidu). This data are collected by keeping track of the locations of smartphones that use the applications from Baidu and can be seen as reflecting the trend of the number of people moving out since Baidu is set as the default search engine for most people in China. Another data set used in this paper reflects the percentage of people moving to the corresponding destinations from their starting points and is also available on [29]. The traveling volume is estimated by combining these two data sets.

Data in Europe include the number of confirmed cases, recovered cases, and fatalities in the following 11 countries from May 1st to August 31st: Denmark, Finland, Norway, Austria, Germany, Switzerland, Italy, Spain, Belgium, France, and Ireland. The data are retrieved from [30, 12]. The corresponding total population in each country is from [31]. However, to our best knowledge, the transportation data between these countries in Europe needed in this model are not publicly available. Thus, the models are simplified when analyzing the European data, and the details can be seen in Section 4.3.

4.2 Results for COVID-19 Data in China

4.2.1 Data Description

After the filtration process described in Appendix B, $n = 21$ provinces or municipalities are taken into consideration, which are Anhui, Beijing, Fujian, Gansu, Guangdong, Guangxi, Hebei, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Liaoning, Ningxia, Shandong, Shanxi, Shanxi, Shanghai, Sichuan, Zhejiang, and Chongqing.

We set $T = 31$ (from January 10th to February 10th) and $T_{th} = 27$ (February 5st). The data from $T_{th,1} = 11$ (January 21st) to the T_{th}-th day (February 5st) are treated as the training set, and the data from $T_{th} + 1$ (February 6th) to the t-th day (February 10th) are treated as the testing set. In addition, from observation, we choose $T_k = 24$ (February 2nd) to be the day when λ_k changes in Anhui, Beijing, Henan and Hubei, and $T_k = 20$ (January 29th) for the other provinces. More details of the data set can be found in Appendix B.

In addition, the data from Baidu Qianxi might not be the exact traveling volumes between municipalities and provinces. We assume that the actual traveling volume from one starting point to one destination is proportional to the product of numbers of people departed from the stating point and the percentage of population travelling from this origin to the destination. The corresponding scaling parameter α also needs to be inferred from the data for all the models.

4.2.2 Models compared

As in Section 3.1.2, the same Models 1-5 are compared for COVID-19 data in China. Recall that Model 5 is the model proposed in this paper and the other four are the baseline models for comparison.

Note that for Model 5 with prior distribution based on graph Laplacian, the provinces are chosen to be divided into two groups, which are Hubei and other provinces except Hubei, due to the overwhelmed medical resources in Wuhan at the early stage of the pandemic ([32]). For simplicity of comparison, we choose $2\mu_0 = \mu_1 = \mu_2$. The parameter σ in (2.7) is chosen to be 1 in this case. Then for Model 5, by (2.9), $\widetilde{\Theta} = \{E_k(0), H_k(0), \lambda_{k,1}, \lambda_{k,2}\}_{k=1}^n$ and scaling parameter α are estimated by the optimization problem in (4.1):

$$\begin{aligned}\widetilde{\Theta}^*, \alpha^* = \arg\min(&-\log p(y_{1:T_{th}}|\widetilde{\Theta}, \alpha) + \mu_0 \sum_{i\in D_1, j\in D_2} w_{ij}\left((\lambda_{i,1} - \lambda_{j,1})^2 + (\lambda_{i,2} - \lambda_{j,2})^2\right) \\ &+ \frac{1}{2}\mu_1 \sum_{i,j\in D_1 \text{ or } D_2} w_{ij}\left((\lambda_{i,1} - \lambda_{j,1})^2 + (\lambda_{i,2} - \lambda_{j,2})^2\right) + \sigma\left(\|\lambda^{(\mathbf{1})}\|_2^2 + \|\lambda^{(\mathbf{2})}\|_2^2\right)). \end{aligned} \tag{4.1}$$

4.2.3 Results of Trajectory Comparison

Figure 7 shows the true and predicted trajectories for newly confirmed cases in Hubei. Note that in Figure 7,

- The red line shows the true trajectory.
- The blue line shows the predicted deterministic trajectories obtained by running (2.3) with Θ^* (inferred using (4.1) with $\mu_0 = 10^{-3.5}$).
- The blue scatter plots show 100 stochastic trajectories with Θ^*.
- The red scatter plots show 100 stochastic trajectories sampled from the posterior distribution of $\widetilde{\Theta}$ (which is estimated by 5×10^5 MCMC iterations staring from $\widetilde{\Theta}^*$).

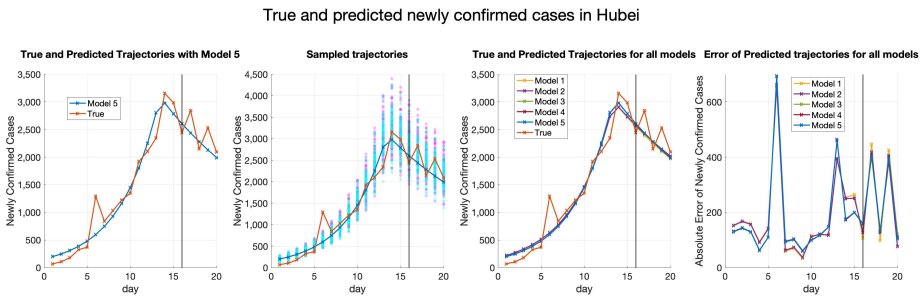

Figure 7: True and fitted trajectories in Hubei. The red line shows the true trajectory, the blue line shows the predicted deterministic trajectory with Θ^* ($\widetilde{\Theta}^*$ inferred using (4.1) with $\mu_0 = 10^{-3.5}$), blue scatter plots show 100 stochastic trajectories with Θ^*, and red scatter plots show 100 stochastic trajectories sampled from the posterior distribution of $\widetilde{\Theta}$ (which is estimated by 5×10^5 MCMC iterations staring from $\widetilde{\Theta}^*$). The vertical lines show the threshold $T_{th} = 27$ for the training data and testing data.

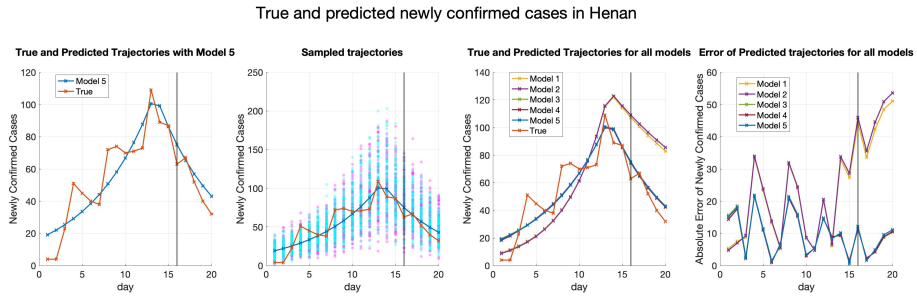

Figure 8: True and fitted trajectories in Henan. The red line shows the true trajectory, the blue line shows the predicted deterministic trajectory with Θ^* ($\widetilde{\Theta}^*$ inferred using (4.1) with $\mu_0 = 10^{-3.5}$), blue scatter plots show 100 stochastic trajectories with Θ^*, and red scatter plots show 100 stochastic trajectories sampled from the posterior distribution of $\widetilde{\Theta}$ (which is estimated by 5×10^5 MCMC iterations staring from $\widetilde{\Theta}^*$). The vertical lines show the threshold $T_{th} = 27$ for the training data and testing data.

The deterministic trajectories obtained by the other four models and the corresponding absolute errors are also plotted for better comparison. It can be seen that predicted deterministic trajectories generally capture the trend of the true trajectory. Moreover, sampling trajectories from the posterior distribution of $\widetilde{\Theta}$ generates more randomness than sampling trajectories with Θ^*. Besides, it can be seen from Figure 7 that the performances of all the five models are similar for Hubei. This may be explained by the fact that the cases in Hubei outnumber those in other provinces or municipalities, which makes all the models tend to fit the trajectory of Hubei best.

Similarly, Figure 8 shows the predicted trajectories in Henan. As can be seen in Figure 8, unlike Hubei, heterogeneity of parameters helps the model fit and predict the trajectory of Henan in a more significant manner. When the transmission parameters in the two periods are forced to be the same in all provinces, the increment of the newly confirmed cases of the predicted trajectory is faster than the trend shown in the true trajectory in the first period. Meanwhile, the decrement of the newly confirmed cases of the fitted deterministic trajectory in the second period is also slower than the trend reflected in the true trajectory. Thus, heterogeneity helps improve the performance of fitting and predicting.

Moreover, from Figure 9 which shows the fitted trajectories in Anhui, it can still be seen that hetero-

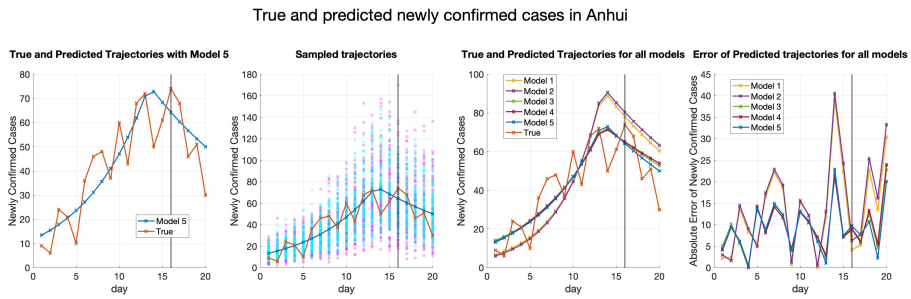

Figure 9: True and fitted trajectories in Anhui. The red line shows the true trajectory, the blue line shows the predicted deterministic trajectory with Θ^* ($\widetilde{\Theta}^*$ inferred using (4.1) with $\mu_0 = 10^{-3.5}$), blue scatter plots show 100 stochastic trajectories with Θ^*, and red scatter plots show 100 stochastic trajectories sampled from the posterior distribution of $\widetilde{\Theta}$ (which is estimated by 5×10^5 MCMC iterations staring from $\widetilde{\Theta}^*$). The vertical lines show the threshold $T_{th} = 27$ for the training data and testing data.

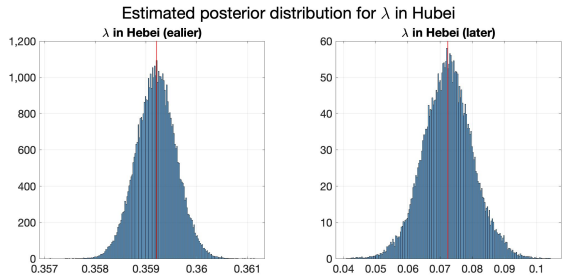

Figure 10: Estimated posterior distribution of λ in Hubei with $\mu_1 = 10^{-3.5}$. The vertical red lines represent the values of corresponding θ^*.

geneity helps the models predict the testing trajectory better. Furthermore, introducing correlation helps Model 5 predict the trajectory marginally better than the other models.

4.2.4 Results of Parameter Estimation

To illustrate the estimation of the posterior of parameters, Figure 10 shows the estimated posterior distributions of $\lambda_{k,1}$ and $\lambda_{k,2}$ in Hubei using Model 5, in which the red lines show the inferred $\lambda_{k,1}$ and $\lambda_{k,2}$ with $\mu_1 = 10^{-3.5}$, and Table 6 shows the inferred $\lambda_{k,1}$ and $\lambda_{k,2}$ with the mean and standard deviation of their estimated posterior distributions from MCMC.

By comparing the transmission rates in Hubei in the two periods, it can be seen that the transmission rate decreases greatly after February 2nd, which indicates that the containment measures adopted in Hubei against the COVID-19 outbreak are effective.

	θ^*	mean of estimated posterior	standard deviation of estimated posterior
$\lambda_{Hubei,1}$	0.3592	0.3592	3.919×10^{-4}
$\lambda_{Hubei,2}$	0.0724	0.0725	0.0078

Table 6: Estimated paramters in Hubei. θ^* is inferred using (4.1) with $\mu_0 = 10^{-3.5}$. The posterior distribution is estimated by 5×10^5 MCMC iterations.

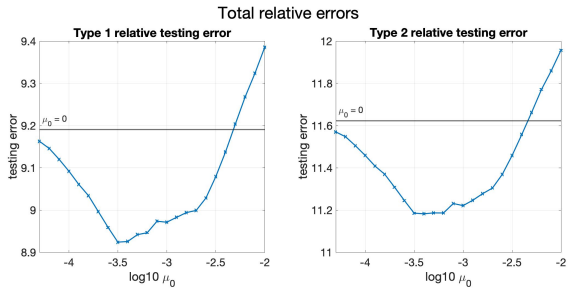

Figure 11: TestRelErr$^{(1)}$ and TestRelErr$^{(2)}$ against μ_0 for real-world data in China. The horizontal green lines show the values of the mean of the testing errors when $\mu_0 = 0$.

4.2.5 Further Model Evaluation

Table 7 shows the training errors and testing errors of the five models computed in the way described in Appendix A. As can be seen from Table 7, both heterogeneity of parameters and transportation between provinces help reduce the training errors and the testing errors as the simulated data case. The utilization of graph Laplacian prior helps further reduce the testing errors. Note that the testing errors of Model 4 are close to those of Model 3, which might be explained as after the traveling restrictions are imposed on January 23rd ([33]), transportation no longer has much influence on the spread of the epidemic.

Model	M	H	GL prior	TrainRelErr$^{(1)}$	TestRelErr$^{(1)}$	TrainRelMSE$^{(1)}$	TestRelMSE$^{(1)}$
1	×	×	×	6.619	9.987	12.232	12.621
2	✓	×	×	6.577	9.807	12.219	12.483
3	×	✓	×	5.652	9.340	11.080	11.895
4	✓	✓	×	5.613	9.191	11.003	11.709
5	✓	✓	✓	5.662	8.924	11.191	11.266

Table 7: Training and testing error for real-world data in China of five models. The first three columns indicate whether the model permits **migration** between provinces, the **heterogeneity** of parameters and the **Graph Laplacian prior**, respectively. $\mu_0 = 10^{-3.5}$ in Model 5 for inference of parameters as defined in (4.1).

Figure 11 below plots TestRelErr$^{(1)}$ and TestRelErr$^{(2)}$ against varying μ_0, and Figure 19 in Appendix C plots TestRelMSE$^{(1)}$ and TestRelMSE$^{(2)}$ against varying μ_0. From Figure 11 and Figure 19, it can be observed that Model 5 first underfits and then overfits as μ_0 decreases.

4.3 Results for COVID-19 Data in Europe

4.3.1 Data Description

11 countries in Europe are considered in the experiments, which are Denmark, Finland, Norway, Austria, Germany, Switzerland, Italy, Spain, Belgium, France, and Ireland.

Since the original data oscillate much, especially over the weekends, the data are first smoothed by taking a simple moving average (which treats the unweighted average of data in 7 days around the i-th day as the new data point for the i-th day). After the smoothing, $T = 116$. Additionally, from the observation,

$T_k = 50$ is chosen to be the day when λ_k changes for all countries except Norway, Italy and France. In addition, $T_{Nowway} = 75$, $T_{Italy} = 85$ and $T_{France} = 60$. Furthermore, we set $T_{th} = 95$.

4.3.2 Models compared

For COVID-19 data in Europe, only three models detailed below are compared, which allow heterogeneity of parameters but do not include transportation. The transportation data needed in this model are not available to the best of our knowledge. However, the impact of transportation is expected to be less significant due to travel restrictions [33].

1'. Model with uniform prior distribution, without heterogeneity or migration.

2'. Model with uniform prior distribution, with heterogeneity but without migration.

3'. Model with prior distribution based on graph Laplacian, with heterogeneity but no migration.

For Model 3', 11 countries are partitioned into $d = 4$ groups:

- Northern Europe: Denmark, Finland, Norway;
- Central Europe: Austria, Germany, Switzerland;
- Southern Europe: Italy, Spain;
- Western Europe: Belgium, France, Ireland,

Besides, in Model 3', we assume that $\mu_1 = \cdots = \mu_d = \mu_{within}$, $\mu_0 = \mu_{between}$, and $\mu_{within} = 10\mu_{between}$ for simplicity. The parameter σ in the prior distribution (2.7) is still taken to be 1. Then by (2.9), for Model 3', $\widetilde{\Theta} = \{E_k(0), H_k(0), \lambda_{k,1}, \lambda_{k,2}\}_{k=1}^n$ are estimated by the optimization problem in (4.2):

$$\begin{aligned}\widetilde{\Theta}^* = &\arg\min(-\log p(y_{1:T_{th}}|\widetilde{\Theta}) + \mu_{between} w \sum_{i\in D_k, j\in D_l, k\neq l} \left((\lambda_{i,1}-\lambda_{j,1})^2 + (\lambda_{i,2}-\lambda_{j,2})^2\right) \\ &+ \frac{1}{2}\mu_{within} w \sum_{i,j\in D_k, k\in\{1,\ldots,d\}} \left((\lambda_{i,1}-\lambda_{j,1})^2 + (\lambda_{i,2}-\lambda_{j,2})^2\right) + \sigma\left(\|\lambda^{(\mathbf{1})}\|_2^2 + \|\lambda^{(\mathbf{2})}\|_2^2\right)), \qquad (4.2)\end{aligned}$$

where $w = 10^5$ is a constant scaling.

4.3.3 Results of Trajectory Comparison

Figure 12, 13 and 14 present the true and predicted trajectories in Austria, Germany, and Italy, respectively. As can be seen from Figure 12, 13 and 14, heterogeneity of transmission parameters in Model 2' and Model 3' helps improve the performance of fitting and generalization. Furthermore, utilization of correlation between countries in Model 3' further improves the prediction of the trajectories.

4.3.4 Results of Parameter Estimation

Same as before, Figure 15 shows the estimated posterior distributions of $\lambda_{k,1}$ and $\lambda_{k,2}$ in Italy using Model 3', in which the red lines show the inferred $\lambda_{k,1}$ and $\lambda_{k,2}$ with $\mu_{between} = 1$. Table 8 shows the inferred $\lambda_{k,1}$ and $\lambda_{k,2}$ and the mean and standard deviation of their estimated posterior distribution from MCMC.

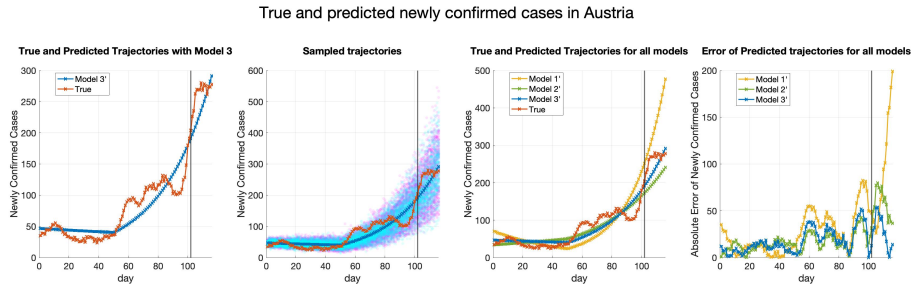

Figure 12: True and fitted trajectories in Austria. The red line shows the true trajectory, the blue line shows the predicted deterministic trajectory with Θ^* ($\widetilde{\Theta}^*$ inferred using (4.2) with $\mu_{between} = 10^{-1}$), blue scatter plots show 100 stochastic trajectories with Θ^*, and red scatter plots show 100 stochastic trajectories sampled from the posterior distribution of $\widetilde{\Theta}$ (which is estimated by 5×10^5 MCMC iterations staring from $\widetilde{\Theta}^*$). The vertical lines show the threshold $T_{th} = 95$ for the training data and testing data.

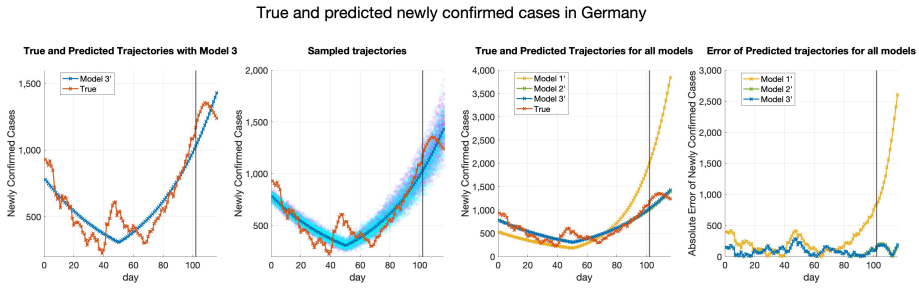

Figure 13: True and fitted trajectories in Germany. The red line shows the true trajectory, the blue line shows the predicted deterministic trajectory with Θ^* ($\widetilde{\Theta}^*$ inferred using (4.2) with $\mu_{between} = 10^{-1}$), blue scatter plots show 100 stochastic trajectories with Θ^*, and red scatter plots show 100 stochastic trajectories sampled from the posterior distribution of $\widetilde{\Theta}$ (which is estimated by 5×10^5 MCMC iterations staring from $\widetilde{\Theta}^*$). The vertical lines show the threshold $T_{th} = 95$ for the training data and testing data.

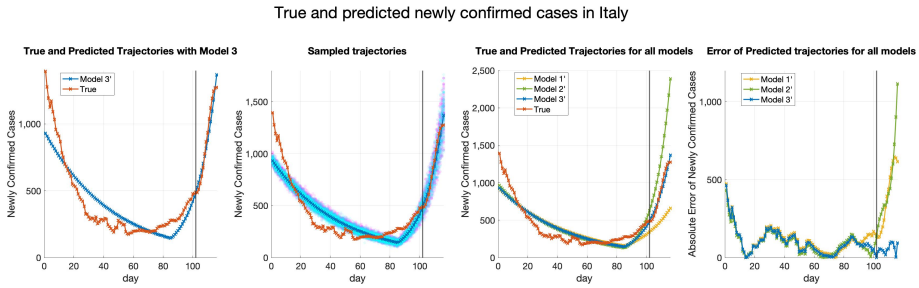

Figure 14: True and fitted trajectories in Italy. The red line shows the true trajectory, the blue line shows the predicted deterministic trajectory with Θ^* ($\widetilde{\Theta}^*$ inferred using (4.2) with $\mu_{between} = 10^{-1}$), blue scatter plots show 100 stochastic trajectories with Θ^*, and red scatter plots show 100 stochastic trajectories sampled from the posterior distribution of $\widetilde{\Theta}$ (which is estimated by 5×10^5 MCMC iterations staring from $\widetilde{\Theta}^*$). The vertical lines show the threshold $T_{th} = 95$ for the training data and testing data.

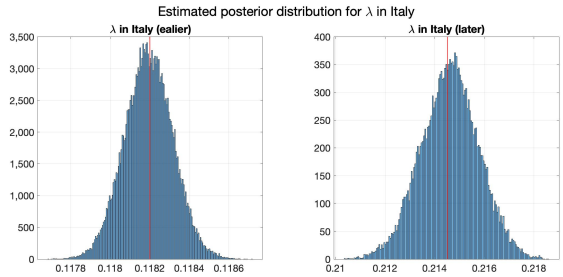

Figure 15: Estimated posterior distributions of λ in Italy with $\mu_{between} = 10^{-1}$. The vertical red lines represent the values of corresponding θ^*.

	θ^*	mean of estimated posterior	standard deviation of estimated posterior
$\lambda_{Italy,1}$	0.1182	0.1182	1.2231×10^{-4}
$\lambda_{Italy,2}$	0.2145	0.2146	0.0012

Table 8: Estimated paramters in Italy. θ^* is inferred using (4.2) with $\mu_{between} = 10^{-1}$. The posterior distribution is estimated by 5×10^5 MCMC iterations.

4.3.5 Further Model Evaluation

Table 9 shows training and testing errors of the three models. It can be seen from Table 9 that heterogeneity of transmission parameters helps reduce both training and testing errors. We can also see that adding Graph Laplacian prior properly further reduces the testing errors greatly without increasing training errors much.

Model	M	H	GL prior	TrainRelErr$^{(1)}$	TestRelErr$^{(1)}$	TrainRelMSE$^{(1)}$	TestRelMSE$^{(1)}$
1'	×	×	×	0.980	6.173	2.419	6.755
2'	×	✓	×	0.535	5.843	1.361	6.607
3'	×	✓	✓	0.616	3.558	1.587	4.145

Table 9: Training and testing error for real-world data in Europe of five models. The first three columns indicate whether the model permits **migration** between provinces, the **heterogeneity** of parameters and the **Graph Laplacian prior**, respectively. $\mu_0 = 10^{-1}$ in Model 3' for inference of parameters as defined in (4.1).

Figure 16 below and Figure 20 in Appendix C plot TestRelErr$^{(1)}$, TestRelErr$^{(2)}$, TestRelMSE$^{(1)}$ and TestRelMSE$^{(2)}$ against varying $\mu_{between}$, from which one may see the same trend of prediction performance as $\mu_{between}$ varies.

5 Discussion

In this paper, we propose a stochastic dynamic model inspired by [1], considering the inter-district transportation and both spatially and temporally heterogeneous transmission parameters, which can model the ongoing and lasting spread of the epidemic in multiple districts. Based on this model, we also introduce a two-step procedure for estimating the parameters, which maximizes over posterior distribution with a

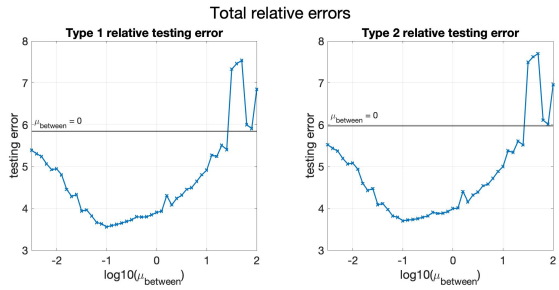

Figure 16: TestRelErr$^{(1)}$ and TestRelErr$^{(2)}$ against $\mu_{between}$ for real-world data in Europe. The horizontal green lines show the values of the mean of the testing errors when $\mu_{between} = 0$.

prior distribution that utilizes graph Laplacian regularization to address the correlation between districts. Experiments on both simulated and real-world data show that compared with the baseline models, this new model with transportation, heterogeneity of parameters, and usage of the correlation in parameter inference improves the performance of fitting and generalization.

We acknowledge that there are limitations in this work. First, the stochastic dynamic model proposed in this paper might not be comprehensive enough to characterize the spread of COVID-19 in reality. For example, the model does not account for the asymptomatic virus carriers and the infectious incubation period. Second, the proposed model does not consider the heterogeneity of risk of being infected in different populations, for example, populations with different ages, jobs, or health conditions. Third, The proposed model needs further modification considering the large-scale COVID-19 vaccination. At last, the current method does not provide the guidance to determine how intensively we would like to stress the correlation between districts.

The current methods have possible extensions in the following several directions in the future study. First, the dynamic model is to be refined to be closer to reality, for example, taking the asymptomatic virus carriers and medical tracking into account and allowing the change of the transmission parameters as time increases to be more realistic. Second, the model can be modified to accommodate the heterogeneity of various populations, including different age groups and vaccination statuses. For example, the compartments S, E, H, R are to be further divided into subgroups, and the parameters such as transmission rates and mortality rates are allowed to be different; meanwhile, the dynamic model is also to be adjusted accordingly. At last, the method could be further used to assess the influence of containing measures taken by different countries, for example, through adjusting the traveling volume to make it different from the actual transportation data and then analyzing the corresponding changes in newly confirmed cases.

Acknowledgement

The authors thank Dr. Huaiyu Tian for helpful discussion. Yuan Zhang and Xiao-Hua Zhou are funded by National Natural Science Foundation of China grant 8204100362 and The Bill & Melinda Gates Foundation (INV-016826). Yixuan Tan and Xiuyuan Cheng are funded by NSF (DMS-2007040) and the Alfred P. Sloan Foundation.

References

[1] Balcan, D. *et al.* Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of computational science* **1**, 132–145, DOI: https://doi.org/10.1016/j.jocs.2010.07.002 (2010).

[2] World Health Organization (WHO). https://covid19.who.int.

[3] Gomes, M. F. *et al.* Assessing the international spreading risk associated with the 2014 west african ebola outbreak. *PLoS currents* **6**, DOI: https://dx.doi.org/10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5 (2014).

[4] Pastore-Piontti, A. *et al.* Real-time assessment of the international spreading risk associated with the 2014 west african ebola outbreak. In *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases*, 39–56, DOI: https://doi.org/10.1007/978-3-319-40413-4_4 (Springer, 2016).

[5] Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* **368**, 395–400, DOI: https://doi.org/10.1126/science.aba9757 (2020).

[6] Colizza, V. *et al.* Estimate of novel influenza a/h1n1 cases in mexico at the early stage of the pandemic with a spatially structured epidemic model. *PLoS currents* **1**, RRN1129–RRN1129, DOI: https://doi.org/10.1371/currents.rrn1129 (2009).

[7] Balcan, D. *et al.* Modeling the critical care demand and antibiotics resources needed during the fall 2009 wave of influenza a(h1n1) pandemic. *PLoS currents* **1**, RRN1133–RRN1133, DOI: https://dx.doi.org/10.1371/currents.RRN1133 (2009).

[8] Bajardi, P. *et al.* Modeling vaccination campaigns and the fall/winter 2009 activity of the new a(h1n1) influenza in the northern hemisphere. *Emerging health threats journal* **2**, e11–e11, DOI: https://doi.org/10.3402/ehtj.v2i0.7093 (2009;2008;2010;).

[9] Tizzoni, M. *et al.* Real-time numerical forecast of global epidemic spreading: Case study of 2009 a/h1n1pdm. *BMC medicine* **10**, 165–165, DOI: https://doi.org/10.1186/1741-7015-10-165 (2012).

[10] Poletto, C. *et al.* Assessment of the middle east respiratory syndrome coronavirus (mers-cov) epidemic in the middle east and risk of international spread using a novel maximum likelihood analysis approach. *Eurosurveillance* **19**, 20824, DOI: https://doi.org/10.2807/1560-7917.es2014.19.23.20824 (2014).

[11] Balcan, D. *et al.* Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility. *BMC medicine* **7**, 1–12, DOI: https://doi.org/10.1186/1741-7015-7-45 (2009).

[12] Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases* **20**, 533–534, DOI: https://doi.org/10.1016/S1473-3099(20)30120-1 (2020).

[13] Zhang, Q. *et al.* Forecasting seasonal influenza fusing digital indicators and a mechanistic disease model. In *Proceedings of the 26th international conference on world wide web*, 311–319, DOI: https://doi.org/10.1145/3038912.3052678 (2017).

[14] Vespignani, A. Predicting the behavior of techno-social systems. *Science* **325**, 425–428, DOI: https://doi.org/10.1126/science.1171990 (2009).

[15] Donnat, C. & Holmes, S. Modeling the heterogeneity in covid-19's reproductive number and its impact on predictive scenarios. *Journal of Applied Statistics* 1–29, DOI: https://doi.org/10.1080/02664763.2021.1941806 (2021).

[16] Tkachenko, A. V. *et al.* Time-dependent heterogeneity leads to transient suppression of the covid-19 epidemic, not herd immunity. *Proceedings of the National Academy of Sciences* **118**, DOI: https://doi.org/10.1073/pnas.2015972118 (2021).

[17] Moreno, Y., Pastor-Satorras, R. & Vespignani, A. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems* **26**, 521–529, DOI: https://doi.org/10.1140/epjb/e20020122 (2002).

[18] Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* **115**, 700–721, DOI: https://doi.org/10.1098/rspa.1927.0118 (1927).

[19] Hilton, J. & Keeling, M. J. Estimation of country-level basic reproductive ratios for novel coronavirus (sars-cov-2/covid-19) using synthetic contact matrices. *PLoS computational biology* **16**, e1008031, DOI: https://doi.org/10.1371/journal.pcbi.1008031 (2020).

[20] Szapudi, I. Heterogeneity in sir epidemics modeling: superspreaders and herd immunity. *Applied Network Science* **5**, 1–12, DOI: https://doi.org/10.1007/s41109-020-00336-5 (2020).

[21] Volpert, V., Banerjee, M. & Sharma, S. Epidemic progression and vaccination in a heterogeneous population. application to the covid-19 epidemic. *Ecological Complexity* 100940, DOI: https://doi.org/10.1016/j.ecocom.2021.100940 (2021).

[22] Hou, X. *et al.* Intracounty modeling of covid-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences* **118**, DOI: https://doi.org/10.1073/pnas.2020524118 (2021).

[23] Chen, S., Li, Q., Gao, S., Kang, Y. & Shi, X. State-specific projection of covid-19 infection in the united states and evaluation of three major control measures. *Scientific reports* **10**, 1–9, DOI: https://doi.org/10.1038/s41598-020-80044-3 (2020).

[24] Zhang, Y. *et al.* Prediction of the covid-19 outbreak based on a realistic stochastic model. *medRxiv* DOI: https://doi.org/10.1101/2020.03.10.20033803 (2020).

[25] Kurtz, T. G. Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of applied Probability* **7**, 49–58, DOI: https://doi.org/10.2307/3212147 (1970).

[26] Kurtz, T. G. Limit theorems for sequences of jump markov processes approximating ordinary differential processes. *Journal of Applied Probability* **8**, 344–356, DOI: https://doi.org/10.2307/3211904 (1971).

[27] He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine* **26**, 672–675, DOI: https://doi.org/10.1038/s41591-020-1016-z (2020).

[28] National Bureau of Statistics of China. Annual data by province (2019). http://www.stats.gov.cn/english/Statisticaldata/AnnualData/.

[29] Baidu Qianxi. https://qianxi.baidu.com.

[30] Johns Hopkins University. Covid-19 data repository by the center for systems science and engineering (csse) at johns hopkins university (2020). https://github.com/CSSEGISandData/COVID-19.

[31] Wikipedia. List of european countries by population. https://en.wikipedia.org/wiki/List_of_European_countries_by_population.

[32] Information Office of Hubei Provincial People's Government. Prevention and control of pneumonia outbreak of new coronary virus infection (2020). https://www.hubei.gov.cn/hbfb/xwfbh/202002/t20200210_2023490.shtml.

[33] Blavatnik School of Government. Covid-19 government response tracker. https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker.

Appendix A Model Evaluation Metric

Measurements are proposed in this section to evaluate the performance of fitting and generalization of the aforementioned models in the experiments section.

To be more specific, the performance of fitting is measured on the training data $\{(\widehat{\Delta C_a})_k^{[Tr]}(i), k = 1, \ldots, n\}_{i=1}^{T_{th}}$ up to some threshold T_{th}, and the performance of generalization is evaluated on the testing data $\{(\widehat{\Delta C_a})_k^{[Te]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$. Note that T_{th} should not be earlier than the maximum of T_k , which is the day that λ_k changes in the k-th region, $\forall k = 1, \ldots, n$.

For both simulated and real-world data, the fitting performance is assessed by comparing the training data $\{(\widehat{\Delta C_a})_k^{[Tr]}(i), k = 1, \ldots, n\}_{i=1}^{T_{th}}$ with the fitted deterministic trajectory $\{(\Delta C_a)_k^{[Tr,\Theta]}(i), k = 1, \ldots, n\}_{i=1}^{T_{th}}$, which is obtained by integrating ODE system (2.3) whose parameters $\Theta = \{E_k(0), H_k(0), \lambda_k, \gamma_k\}_{k=1}^{n}$ are inferred by the corresponding model.

Similarly, the generalization performance is assessed by comparing the testing data $\{(\widehat{\Delta C_a})_k^{[Te]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$ with the predicted deterministic trajectory $\{(\Delta C_a)_k^{[Te,\Theta]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$ obtained by integrating (2.3) with parameters Θ. Although the deterministic trajectory is still computed by ODE system (2.3), its initialization is the number of individuals in compartments on the T_{th}-th day of the testing data, i.e., $\{S_k(T_{th}), E_k(T_{th}), H_k(T_{th}), R_k(T_{th}), k = 1, \ldots, n\}$. Then, the trajectory $\{(\Delta C_a)_k^{[Te,\Theta]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$ is obtained by running (2.3) for $T - T_{th}$ days with the transmission parameters $\{\lambda_k\}_{k=1}^{n}$.

Simulated data

For simulated data, two random trajectories are sampled according to the stochastic model with prefixed parameters in one replica. Data from the first T_{th} days of one trajectory are treated as the training data, and data from the last $T - T_{th}$ days of the other trajectory are treated as the testing data. This guarantees that the training and testing data are independent.

Besides, the prestored $\{S_k(T_{th}), E_k(T_{th}), H_k(T_{th}), R_k(T_{th})\}$ of the testing trajectory are used as the initialization of (2.3) to obtain the deterministic trajectory $\{(\Delta C_a)_k^{[Te,\Theta]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$ as described above.

real-world data

For real-world data, only one observable trajectory is available, thus the training data and the testing data are the two parts of the true trajectory split by the T_{th}-th day.

In contrast to the case for simulated data, for real-world data, only $\{(\widehat{C_a})_k(T_{th}), (\widehat{R_a})_k(T_{th})\}$ are available, while $\{E_k(T_{th}), H_k(T_{th}), R_k(T_{th})\}$ are not known. Therefore, trajectories starting from the $T_{th}+1$-th day can not be computed directly. Instead, a deterministic trajectory for all the T days is computed with the fitted parameters Θ, and then the data for the first T_{th} days and for the last $T - T_{th}$ days are treated as $\{(\Delta C_a)_k^{[Tr,\Theta]}(i), k = 1, \ldots, n\}_{i=1}^{T_{th}}$ and $\{(\Delta C_a)_k^{[Te,\Theta]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$, respectively.

Based on the explanations above, the training errors and testing errors can be defined. Details are listed below.

Remind that the ground truth training data and testing data are $\{(\widehat{\Delta C_a})_k^{[Tr]}(i), k = 1, \ldots, n\}_{i=1}^{T_{th}}$ and $\{(\widehat{\Delta C_a})_k^{[Te]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$, respectively. In addition, the fitted deterministic training trajectories and predicted deterministic testing trajectories with parameters Θ are denoted as $\{(\Delta C_a)_k^{[Tr,\Theta]}(i), k = 1, \ldots, n\}_{i=1}^{T_{th}}$ and $\{(\Delta C_a)_k^{[Te,\Theta]}(i), k = 1, \ldots, n\}_{i=T_{th}+1}^{T}$, respectively. The details of obtaining training data, testing data along with fitted trajectories for both simulated data and real-world data have been explained above.

Based on the ground truth trajectories and the fitted ones, define

$$\text{TrainAbsErr}_k(i) = |(\widehat{\Delta C_a})_k^{[Tr]}(i) - (\Delta C_a)_k^{[Tr,\Theta]}(i)|, \tag{A.1}$$

$$\text{TrainRelErr}_k(i) = \frac{|(\widehat{\Delta C_a})_k^{[Tr]}(i) - (\Delta C_a)_k^{[Tr,\Theta]}(i)|}{(\widehat{\Delta C_a})_k^{[Tr]}(i)}, \tag{A.2}$$

which denote the absolute error and relative error for the training data in the k-th region on the i-th day, respectively.

With the errors defined for each day, the total errors for the k-th region are the weighted sums of $\{\text{TrainAbsErr}_k(i)\}_{i=1}^{T_{th}}$ and $\{\text{TrainRelErr}_k(i)\}_{i=1}^{T_{th}}$. There exist two types of reasonable weights for weighting the errors. The first type is defined as

$$W_{k,train}^{(1)}(i) = \frac{(\widehat{\Delta C_a})_k^{[Tr]}(i)}{\sum_{i'=1}^{T_{th}} (\widehat{\Delta C_a})_k^{[Tr]}(i')}, \tag{A.3}$$

which share with same trend of increasing or decreasing as the newly confirmed cases in the k-th region.

Moreover, the second type of weights is the simple average of errors for all the T days, that is to say,

$$W_{k,train}^{(2)}(i) = \frac{1}{T_{th}}, \quad \forall k \in \{1, \ldots, n\}, i \in \{1, \ldots, T_{th}\}. \tag{A.4}$$

With the weights defined as (A.3) or (A.4), the total absolute or relative training errors or mean square errors for the k-th region can be correspondingly defined as

$$\text{TrainAbsErr}_k^{(j)} = \sum_{i=1}^{T_{th}} W_{k,train}^{(j)}(i)\text{TrainAbsErr}_k(i), \tag{A.5}$$

$$\text{TrainRelErr}_k^{(j)} = \sum_{i=1}^{T_{th}} W_{k,train}^{(j)}(i)\text{TrainRelErr}_k(i), \tag{A.6}$$

$$\text{TrainAbsMSE}_k^{(j)} = \left(\sum_{i=1}^{T_{th}} W_{k,train}^{(j)}(i)(\text{TrainAbsErr}_k(i))^2\right)^{1/2}, \tag{A.7}$$

$$\text{TrainRelMSE}_k^{(j)} = \left(\sum_{i=1}^{T_{th}} W_{k,train}^{(j)}(i)(\text{TrainRelErr}_k(i))^2\right)^{1/2}, \tag{A.8}$$

where $j \in \{1, 2\}$. After defining the errors for each region, the total errors are defined as the summation

of the errors of all the regions,

$$\text{TrainAbsErr}^{(j)} = \sum_{k=1}^{n} \text{TrainAbsErr}_k^{(j)}, \tag{A.9}$$

$$\text{TrainRelErr}^{(j)} = \sum_{k=1}^{n} \text{TrainRelErr}_k^{(j)}, \tag{A.10}$$

$$\text{TrainAbsMSE}^{(j)} = \sum_{k=1}^{n} \text{TrainAbsMSE}_k^{(j)}, \tag{A.11}$$

$$\text{TrainRelMSE}^{(j)} = \sum_{k=1}^{n} \text{TrainRelMSE}_k^{(j)}. \tag{A.12}$$

In addition, the testing errors and weights are defined in the same way as above, which are listed below,

- Absolute error and relative error of the k-th region on the i-th day:

$$\text{TestAbsErr}_k(i) = |(\widehat{\Delta C_a})_k^{[Te]}(i) - (\Delta C_a)_k^{[Te,\Theta]}(i)|, \quad \text{TestRelErr}_k(i) = \frac{|(\widehat{\Delta C_a})_k^{[Te]}(i) - (\Delta C_a)_k^{[Te,\Theta]}(i)|}{(\widehat{\Delta C_a})_k^{[Te]}(i)}.$$

- Two kinds of weights:

$$W_{k,test}^{(1)}(i) = \frac{(\widehat{\Delta C_a})_k^{[Te]}(i)}{\sum_{i'=T_{th}+1}^{T} (\widehat{\Delta C_a})_k^{[Te]}(i')}, \quad W_{k,test}^{(2)}(i) = \frac{1}{T - T_{th}}.$$

- Weighted absolute error and relative error of the k-th region:

$$\text{TestAbsErr}_k^{(j)} = \sum_{i=T_{th}+1}^{T} W_{k,test}^{(j)}(i)\text{TestAbsErr}_k(i), \quad \text{TestRelErr}_k^{(j)} = \sum_{i=T_{th}+1}^{T} W_{k,test}^{(j)}(i)\text{TestRelErr}_k(i).$$

- Weighted absolute and relative mean square error of the k-th region:

$$\text{TestAbsMSE}_k^{(j)} = \left(\sum_{i=T_{th}+1}^{T} W_{k,test}^{(j)}(i)(\text{TestAbsErr}_k(i))^2 \right)^{1/2},$$

$$\text{TestRelMSE}_k^{(j)} = \left(\sum_{i=T_{th}+1}^{T} W_{k,test}^{(j)}(i)(\text{TestRelErr}_k(i))^2 \right)^{1/2}.$$

- Total weighted absolute and relative error:

$$\text{TestAbsErr}^{(j)} = \sum_{k=1}^{n} \text{TestAbsErr}_k^{(j)}, \quad \text{TestRelErr}^{(j)} = \sum_{k=1}^{n} \text{TestRelErr}_k^{(j)}.$$

- Total weighted absolute and relative mean square error:

$$\text{TestAbsMSE}^{(j)} = \sum_{k=1}^{n} \text{TestAbsMSE}_k^{(j)}, \quad \text{TestRelMSE}^{(j)} = \sum_{k=1}^{n} \text{TestRelMSE}_k^{(j)}.$$

In the above definitions, $i \in \{1, \ldots, T\}, k \in \{1, \ldots, n\}, j \in \{1, 2\}$.

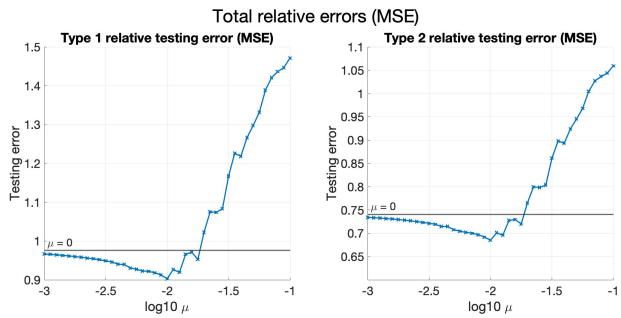

Figure 17: Mean of TestRelMSE$^{(1)}$ and TestRelMSE$^{(2)}$ against μ for simulated data (two provinces). The horizontal green lines show the values of the mean of the testing errors when $\mu = 0$.

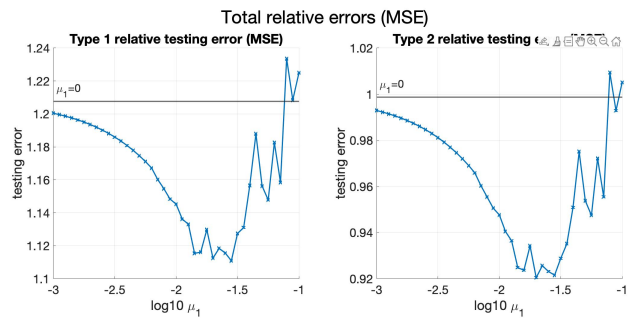

Figure 18: Mean of TestRelMSE$^{(1)}$ and TestRelMSE$^{(2)}$ against μ_1 for simulated data (four provinces). The horizontal green lines show the values of the mean of the testing errors when $\mu_1 = 0$.

Appendix B Preprocessing of real-world data in China

Filtration of Provinces in China

For stability of inference for parameters, only the provinces or municipalities whose accumulated confirmed cases are above 50, the estimated δ_k are not less than 10^{-10} and the change of λ_k occurs no later than February 2nd are considered.

Description of real-world data in China

The COVID-19 data including the number of confirmed cases, recovered cases and fatalities in major provinces and cities of China are from January 21st to March 28th. Moreover, the transportation data are from January 10th to February 10th. Thus, we consider the spread of COVID-19 in China from January 10th to February 10th, and $T = 31$. However, only the COVID-19 after January 21st are available, hence the training set is the data from $T_{th,1} = 11$ (January 21st) to the T_{th}-th day (February 5st), and the testing set is the data from $T_{th} + 1$ (February 6th) to the t-th day (February 10th).

Appendix C Supplementary Figures

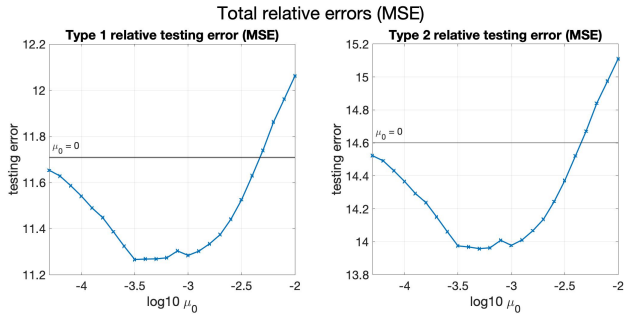

Figure 19: TestRelMSE$^{(1)}$ and TestRelMSE$^{(2)}$ against μ_0 for real-world data in China. The horizontal green lines show the values of the mean of the testing errors when $\mu_0 = 0$.

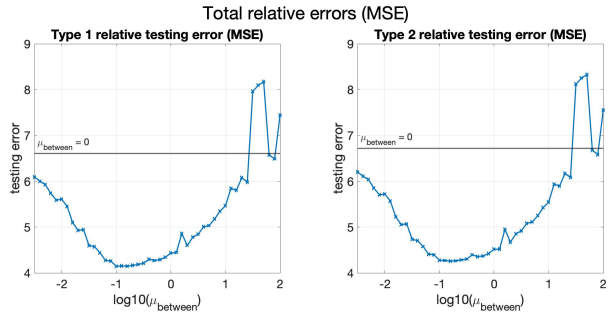

Figure 20: TestRelMSE$^{(1)}$ and TestRelMSE$^{(2)}$ against $\mu_{between}$ for real-world data in Europe. The horizontal green lines show the values of the mean of the testing errors when $\mu_{between} = 0$.